

Medical Students Deciding Their Future: Hospital Edition

Joana Rocha Camargo, Abdullah Al-Alami, Thiago Marcal, Barbara Holguin

IE University

Econ/Data Lab

Rodrigo Alegria Huerta

June 24, 2022

Table of Contents

- Abstract: 3
- Keywords: 4
- Introduction: 5
- Literature Review: 6
- Methodology: 11
- Descriptive Statistics: 14
- Further Analysis: 25
- Discussion: 32
- References: 34
- Appendix: 36

Abstract

One of the most important considerations for hospitals when looking to increase or maintain the number of applications and interested individuals in working at their hospital is their decision factors. Due to looking to expand their operations, Hospital Universitario de La Paz has invited us to investigate this same question. What are the main factors for medical students when choosing a hospital to complete their specialization at? In this report we will collect and analyze data from the workers at Hospital Universitario de La Paz to get an insight into their decision and make recommendations to the hospital for future considerations, improvements or characteristics to maintain into the future as the size of the hospital increases. Through literature, it was found that the three most important factors were reputation, location and variety of case mix. Attempting to confirm or deny these claims, a survey, descriptive statistics, correlations, and further graphical analysis will be done to reach final conclusions.

Keywords

Pearson Correlation

P-value

Quality Training

Case Mix

Aggregate Data

Variation

Normality

Introduction

Graduates of medical schools in Spain who aim to become specialists ought to work in certified hospitals as part of residency programs for a period ranging from three to five years. These hospitals are usually better-equipped, developed and larger compared to Spain's other healthcare institutions. Joining the residency programs requires potential residents to sit for the MIR exam, whose score gets averaged with the previous medical school grade. Thereafter, the weighted averages are used to determine the graduates' ranks during the vacancy selection process such that those with higher ranks get to choose the hospital where they would complete the residency program before those with lower ranks (Machado et al, 2021). With that said, apart from the rank which the graduate receives, there are many factors that contribute to candidates' choice of certain hospitals or residency programs such as training quality, location, and reputation. Therefore, the aim of this report is to determine what specific factors graduates of medical schools take into consideration when choosing the residency programs in Spain. This will be done primarily based on data collected from a survey distributed to program residents of Hospital Universitario de La Paz - one of Madrid's most important healthcare institutions training medical graduates. Thereafter, the collected data will be processed using three analysis tools; descriptive statistics, graphical analysis, and Pearson correlation.

Literature Review

Before digging into the reasons why medical students choose the specializations and hospitals they do, it seemed important to understand their process before making said choices. It was found that after finishing studies it was mandatory to take a national selective exam whose score was later joined with their overall grades during studies to put them in a list based on their final grade (Freire et al, 2015). Typically, students that scored the worst and were on the bottom of the list were left with very few options of specializations and many times had to decide on taking said specialization or dropping out to take the test again the next year (Harris et al, 2016). Closer to the end of the specialization training, the individuals are given qualifications of their actions and quality of their jobs during the years they participated. With this they continue to get a job, many times at the hospital they completed their specialization at. In terms of the expectations of the hospitals, it is important to mention that “health facilities interested in the training of medical specialists need to pass a formal process of accreditation”, meaning that not all hospitals and health facilities are allowed to do so (Freire et al, 2015). While this is very interesting and could be important when viewing individuals' decision factors out of university to complete their specialization, in this study we will not take this into consideration due to the fact that other individuals could have transferred after completing these and their opinions and experiences are also valid. We recommend that in a future study, this path be investigated. We will handle all hospitals as possible options for individuals when it comes to investigating the factors in hospital choice.

Current literature in Spain showed that there are several factors that medical graduates take into consideration when choosing their vacancies for the residency programs. The predominant factor driving candidates' decisions is the training quality the hospital has (Machado et al, 2012). Nonetheless, the study showed that the *perceived* quality varies from candidate to candidate. For instance, some candidates measure the training quality by the

casemix - that is the variety of medical cases treated by the hospitals. On the other hand, others measure it through the workload, or the general care quality the hospital offers. Additionally, external parties such as media, relatives, professors and former students' information about training hospitals were shown to influence candidates' perception- however to a less significant extent (Machado et al, 2012). Similarly, when it comes to the specialization individuals choose, workload, care quality, and influence of others were also proven to be the most important factors in this decision (Kazerooni et al, 2021). With that said, objectifying the factor "training quality" was proven to be problematic. Thus, it makes more sense to consider the casemix, hospital size, as well as the reputation from media and relatives as separate factors.

Furthermore, other factors such as location were important, though relatively insignificant compared to the perceived training quality. For example, candidates were discriminating between hospitals within the same geographical location based on the training quality (Machado et al, 2012). As for pay, public hospitals in Spain provide their residents with similar wages. Therefore, pay is not an important factor compared to training quality and location. Instead, we assume that this factor could only affect the residents' choice of specialization. The argument is that certain specializations are more promising in terms of pay in the long-run. In fact, when analyzing specializations chosen, it was found that the three most desired specializations were cardiology, dermatology and plastic surgery, very high paying jobs (Harris et al, 2016). Thus, to attract more residents, hospitals need to have a high variety of casemix and specializations.

Looking at literature outside of Spain, similar factors were prominent in medical students' choices of specializations and training hospitals/programs. For example, in Canada, medical graduates need to rank residency programs as part of the Canadian Residency Matching Service (CaRMS) which facilitates students' residency choices (Wang et al, 2011).

Having said that, graduating students from six medical schools in Ontario were surveyed to analyze what factors they take into consideration when choosing residency programs. Likewise to Spain, Canadian graduates' responses showed that the "variety of clinical experiences" (aka casemix in Spanish literature) is weighted heavily when choosing the residency program. In parallel, the same importance was given to the residents' and hospital morale-i.e the reputation (Wang et al, 2011). Additionally, other major factors included location; specifically distance to the graduates' relatives' as well as the citylife. As for "financial incentives", the factor was weighted poorly similarly to the situation in Spain (Wang et al, 2011). Moreover, a cross-sectional survey was created in 1999 for medical graduates applying to four residency programs in the US during the National Residency Matching Plan (NRMP) where 1,005 out of 2,820 surveys were answered. With that said, it was found that graduates value program quality, reputation and location as the most important factors influencing their decision of certain residency programs (Aagaard et al, 2005). However, it's important to note that women and minority ethnic groups placed importance on diversity, meaning that recruiters of the residency program ought to hire diverse groups and genders of people (Aagaard et al, 2005). As a matter of fact, in another study where 148 out of 316 surveys were answered, female and underrepresented groups placed higher importance on factors of gender and ethnic diversity to be present in a residency program compared to males (Agawu et al, 2019). Nonetheless, this was still valued less compared to the factors mentioned prior.

Therefore, the predominant factor influencing medical students' choices of residency programs across different countries seems to be the training quality-specifically measured by the casemix or the variety of medical cases as a separate factor. After that, location and the hospital's reputation seem to be important driving factors, though to a less significant extent.

Three Most Important Decision Factors:	High Variety of Case-Mix	Location: Distance to city center and relatives/friends	Reputation
--	--------------------------	---	------------

Table 1 - Table of Most Important Decision Factors

DISTRIBUCIÓN POR TRAMOS DE EDAD DE LOS ASPIRANTES ADJUDICATARIOS DE PLAZA EN LAS PRUEBAS SELECTIVAS DE 2020-21
(Se utiliza el 31-12-2021 como fecha de referencia)

	EDAD														TOTAL
	De 20 a 24		De 25 a 29		De 30 a 34		De 35 a 39		De 40 a 44		De 45 a 49		= > 50 años		
	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	
Medicina	33	0,4%	6.680	83,6%	881	11,0%	234	2,9%	97	1,2%	37	0,5%	25	0,3%	7.987
Farmacia	88	33,1%	165	62,0%	11	4,1%		0,0%		0,0%		0,0%	2	0,8%	266
Enfermería	800	47,5%	613	36,4%	173	10,3%	65	3,9%	26	1,5%	3	0,2%	3	0,2%	1.683
Psicología	45	22,7%	112	56,6%	35	17,7%	5	2,5%		0,0%	1	0,5%		0,0%	198
Biología	15	30,0%	30	60,0%	3	6,0%	2	4,0%		0,0%		0,0%		0,0%	50
Química	8	36,4%	10	45,5%	3	13,6%	1	4,5%		0,0%		0,0%		0,0%	22
Física	12	30,0%	22	55,0%	5	12,5%	1	2,5%		0,0%		0,0%		0,0%	40
TOTAL	1.001	9,8%	7.632	74,5%	1.111	10,8%	308	3,0%	123	1,2%	41	0,4%	30	0,3%	10.246

Table 2 - Specializations by Age Group in Pruebas Selectivas

RESUMEN DE LOS DATOS MAS SIGNIFICATIVOS DE LAS PRUEBAS SELECTIVAS 2020/2021

	MEDICINA		FARMACIA		QUÍMICA		BIOLOGÍA		PSICOLOGÍA		FÍSICA		ENFERMERÍA		TOTALES		
FASE DE ADMISIÓN																	
SOLICITUDES PRESENTADAS		15.166		1.598		289		1.157		4.401		220		7.419		30.250	
Nº ASPIRANTES ADMITIDOS A EXAMEN	MUJERES (Nº y %)	9.219	63,8%	1.191	75,2%	180	63,4%	839	73,9%	3.571	82,3%	84	38,7%	6.468	89,3%	21.552	73,7%
	HOMBRES (Nº y %)	5.233	36,2%	392	24,8%	104	36,6%	296	26,1%	767	17,7%	133	61,3%	772	10,7%	7.697	26,3%
	TOTAL	14.452		1.583		284		1.135		4.338		217		7.240		29.249	
FASE DE SELECCIÓN																	
Nº ASPIRANTES PRESENTADOS A EXAMEN	MUJERES (Nº y %)	8.546	64,1%	1.102	75,0%	143	60,9%	730	73,1%	3.184	82,3%	74	38,1%	5.717	89,6%	19.496	73,6%
	HOMBRES (Nº y %)	4.786	35,9%	367	25,0%	92	39,1%	269	26,9%	687	17,7%	120	61,9%	663	10,4%	6.984	26,4%
	TOTAL	13.332		1.469		235		999		3.871		194		6.380		26.480	
% PRESENTADOS / ADMITIDOS A EXAMEN		92,3%		92,8%		82,7%		88,0%		89,2%		89,4%		88,1%		90,5%	
ASPIR. ELIMINADOS TRAS NOTA DE CORTE		2.527	19,0%	166	11,3%	62	26,4%	144	14,4%	1.203	31,1%	51	26,3%	2.273	35,6%		
ASPIR. CON NÚMERO DE ORDEN		10.805		1.303		173		855		2.668		143		4.107			
RATIO SUPERAN CORTE / PLAZA		1,35		4,88		7,86		17,10		13,47		3,58		2,44			
FASE DE ADJUDICACIÓN																	
PLAZAS OFERTADAS		7.988		267		22		50		198		40		1.683		10.248	
PLAZAS ADJUDICADAS	MUJERES (Nº y %)	5.308	66,5%	189	71,1%	11	50,0%	29	58,0%	163	82,3%	14	35,0%	1.545	91,8%	7.259	70,8%
	HOMBRES (Nº y %)	2.679	33,5%	77	28,9%	11	50,0%	21	42,0%	35	17,7%	26	65,0%	138	8,2%	2.987	29,2%
	TOTAL	7.987		266		22		50		198		40		1.683		10.246	
PLAZAS SIN ADJUDICAR		1		1		0		0		0		0		0			
ULTIMO Nº ORDEN CON PLAZA ADJUDIC. TURNO ORDINARIO // TURNO DISCAPACIDAD		9.845 // 10.729		264 // 1.164		20 // 110		45 // 478		184 // 1.726		41 //		2.702// 4.102			
El cupo de plazas para extranjeros no comunitarios se agotó con el número de orden 5.655																	

Table 3 - Table of Gender in Specialization in Different Stages of the Pruebas Selectivas

Due to the information found in the literature, we became interested in investigating if the factors of quality training, location and hospital's reputation were truly important factors for medical students. Additionally, we wondered if specific factors and characteristics of individuals would also impact these preferences. To answer these questions we looked for data in the Spanish Ministry of Health Website. When doing so, we realized that the data was percentages and aggregate data, meaning, the data did not connect to one person and utilized percentages to describe the number of individuals with those characteristics that partook in something or made a specific choice. For example, the percentage of individuals who applied for medicine or pharmaceutical services in different age groups (Table 1), the percentage of medicine applicants that were women (Table 2) and the grades received on the pruebas selectivas for infirmary. Later, through speaking with individuals from the Spanish Ministry of Health, it was found that the type of data we were looking for was no longer available in order to adjust for new privacy laws. This data was not useful for us because as mentioned before, it was not connected to a specific decision and possible patterns in this decision making process. Although this was good to get to know the question at hand and understand the hospital worker experience, we had to find another way to collect data useful for our project. With this, we produced surveys to send out to the Hospital Universitario de la Paz staff. Nevertheless, the sample size was significantly small given that only 21 residents of Hospital Universitario de la Paz's 350 residents have answered the questions. With that said, it is important to note that conclusions arising from the report will be weak.

Methodology

Participants

The participants in this study are workers at the Hospital Universitario de La Paz in Madrid. They are in various specializations and from differing starting years creating a variety of respondent types. These specializations include hematology, clinical pharmacology, plastic and reconstructive surgery and family doctors. The sample size is that of 21 respondents which is 6% of the hospital's workers, therefore, the conclusions made in this report will be weak and possibly not generalizable to the whole Hospital Universitario de La Paz community. Although this is the case, there is a variety in the respondent types which will allow us to have a wider perspective on this topic. The small number of respondents will not allow us to create cause and effect conclusions for Hospital Universitario de La Paz but instead detect possible relationships and importance. Furthermore, we created a second survey for hospital workers outside of Hospital Universitario de La Paz in an attempt to look for possible patterns or similarities between the individuals from different hospitals and also make the conclusions more generalizable to hospitals in Madrid, but due to not receiving any responses, this idea had to be abandoned.

The Survey

For data collection, the team utilized the information found in literature and the aggregate data from the Spanish Ministry of Health Website to create two surveys which included quantitative and qualitative response questions to obtain the overall perspective of the workers of the hospital and their characteristics. One survey was created for those working at Hospital Universitario de La Paz and another for those from other hospitals in Madrid. With this, we hoped to achieve a more complete perspective of the hospital's staff and find patterns or relationships regarding the decision factors between individuals with

characteristics in common. These questions included “what year did you begin your specialization”, “how would you rate importance of size”, “what was your grade on the pruebas selectivas”, among other quantitative questions. From the aggregate data in the Spanish Ministry of Health Website we found a difference between specialization choice for different age groups and starting years. Furthermore, taking into consideration that since 2019 COVID has been an important part of the world, the decisions made and importance of other factors could have changed. With the questions presented above we hope to discover the difference between individuals of different age groups and possibly how COVID has impacted the decision making process for individuals looking for specializations and hospitals. Moving to the “how would you rate the importance of” questions, through the literature found on this topic, we found that there are many factors that play a role in the decision making process, including location, quality of training (specifically variety of medical cases), and reputation. By asking these questions, we hope to confirm these statements or discover why we got contrasting results. In terms of qualitative questions we asked “what were the 3 most important factors”, gender, university attended, and “how did you gain information about the hospital’s training quality”. To make the comparisons and look for patterns in the quantitative questions we had to ask some qualitative ones, including, gender and university attended. From the aggregate data we found that these two characteristics did have an impact on decision factors and that they were often used as describing factors of the individuals. We once again hoped to confirm these hypotheses and possibly explain patterns found. In regards to the “what were the 3 most important factors” questions, we wanted to give individuals the space to possibly add other factors that we might have not mentioned. In contrast, if that was not the case, we hoped that their response would further support the ones we did add to the survey due to the literature and aggregate data. While analyzing the responses and data gathered through the survey, it was found that very

few individuals did master, PhD or other courses after undergrad and the grades on pruebas selectivas utilized differing scales. Because of this, these two columns of the dataset were removed due to missing values or incompatibility.

Assessments and Measures

The main analysis tools utilized in this study are descriptive statistics, graphical analysis and the Pearson correlation. Descriptive analysis informs data scientists about what the data is talking about and gives the first main insights to the data. This can include giving a summary of the distributions of the variables including mean, standard deviation and possibly 25th and 75th percentiles. Furthermore, it can inform individuals of possible missing values and how many responses they have for each variable, since the data was based on a survey. Graphical analysis includes creating graphs to represent the data and possible relationships between the variables. This can be finding the count, average or other measures of the variables in its different levels or seeing how the values of one variable change depending on another. For example, one can see how location is an important factor for an individual based on age or gender and see if there may be a relationship between said variables. Furthermore, it can aid individuals in finding patterns in the data. As another example, seeing how the importance of pay changes depending on the age of the individual when they started and possibly finding a decrease or increase as the age increases. With this, the pattern can be found and if proven to be significantly significant, used to estimate the importance of pay for other individuals depending on their age. Lastly, the Pearson Correlation evaluates the possible relationships between variables. A correlation of 0.80 or higher represents a high correlation, therefore a strong relationship, a correlation between 0.40 and 0.75 represents a moderate correlation and a correlation below 0.40 represents a weak correlation, meaning there exists no relationship between the variables. To check for a possibly significant

correlation between variables we have used the p-value. A significant correlation means that the correlation found between the variables is not solely by chance but that there truly are other factors that make it true. The alpha, the value that the p-value is compared to to check for significance usually is of 0.10, 0.05 or 0.01. For the p-value to be significant, it has to be smaller than the alpha value. For the purpose of this study we will use the alpha value of 0.01.

Due to the small sample size, having only 21 participants and respondents of the survey, the team was unable to run any other tests to search for significant impacts on the decision of individuals to choose Hospital Universitario de La Paz in Madrid to do their specializations at. In order to work around this and still provide a useful analysis, the team utilized descriptive statistics, the visualization of variables, their relationships and possible interactions to find patterns in decision factors for the Hospital Universitario de la Paz team. Furthermore, to add to the research, the team analyzed the aggregate data found on the Spanish Ministry of Health Website to add to the insights found in the surveys and create a more complete picture of the question at hand and its solution.

Descriptive Statistics

This section aims to provide a detailed overview of the main factors contributing to the decision-making process of students when choosing the hospital to do their specialization. Using the data collected from the survey, we used descriptive statistics to analyze our matter to help us reach a conclusion. Three main methods were used in our analyses, the first being boxplots, useful in revealing how centered the data is and the possible outliers. A second method was the use of histograms which determines the frequency of the variables. Finally, a Correlation Matrix, where the relationship between our variables was tested, showing how strongly or weakly correlated they are. Under those statistics methods, the variables

considered were age beginning, year beginning, importance size, importance location, importance pay and importance review. Below is the summary output of data collected from the surveys.

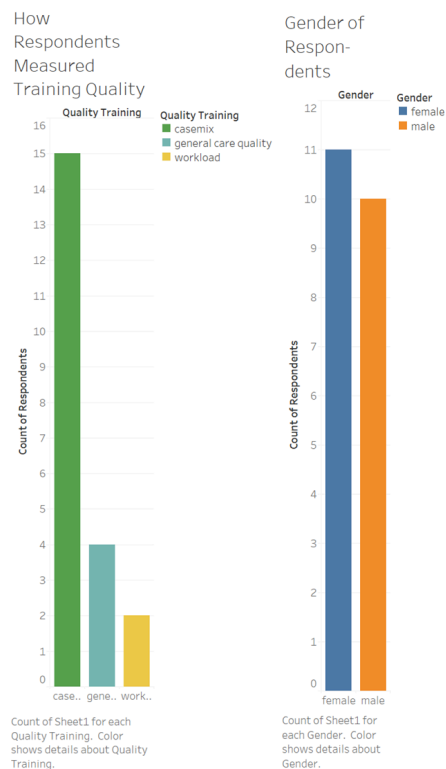
Summary Table of Variables

Variable	Age Beginning	Year Beginning	Importance Size	Importance Location	Importance Pay	Importance Review
Min	247	2017	6	1	0	1
1st Q.	24	2018	8	7	3	5
Median	25	2020	9	9	5	7
Mean	26.52	2020	8.524	7.619	5.19	6.45
Standard Deviation	5.15	1.40	1.17	2.96	3.11	2.25
Variance	26.562	1.957	1.362	8.748	9.662	5.048
3rd Q.	25	2021	9	10	7	8
Max	43	2021	10	10	10	10

Table 4 - Table of Standard Deviations of Variables

Table 4 shows the summary of all the output of the quantitative variables collected, displaying the median, mean, minimum and maximum, standard deviation, variance and the quartile range. This organization of the data information is of great importance as it reveals where the hospital should focus more under each category. For age beginning for instance, it's noticeable that people begin their specialization when they are around 25, usually meaning they are entering the hospital straight after graduating. From the answers collected from the respondents, the majority of them started in the year of 2020. Due to this, we will take into consideration that COVID was already around at this time and because a great majority started in this year, their responses may be different than those beginning before COVID, so the results may not be as effective when it comes to individuals that started

before COVID due to the lack of representation. Next, we can understand from the information given to us about the importance of location that it has a great impact on the students' hospital choice. The average is 8.524 while the minimum was a 6, which is already considerably high. The next variable was the importance of the location of the hospital, which from the data displayed, it also is important for a student's choice. The most important aspect to determine the importance has to be the mean, as you are able to see the center point of the data group. Importance of pay and review on the other hand, display a lower average than the other variables, but still being above 5. The interpretation behind this is that of course students care about the pay and the hospital review, but it is not a major determinant as there are other variables in which they consider to be more important. As for the variance, which informs individuals of the average difference of the points to the mean, shows us that the variable with the most difference in values is age beginning with importance of location and importance of pay behind it. So, for the respondents, the questions that had the biggest range of differences when it came to the decision factors were the three previously mentioned, meaning that for pay and location, individuals' position on this topic and its importance to them differed the most between them.



Graph 1 and 2 - Qualitative Variables Graphs

From graphs 1 and 2 we can see an overview of the qualitative variables. Graph 1 shows us the frequency of gender from the respondents, from which we see that there are more men, but there's a significant equilibrium between genders. From graph 2, we can learn how students measure training quality, casemix being the most important by far.

Boxplots

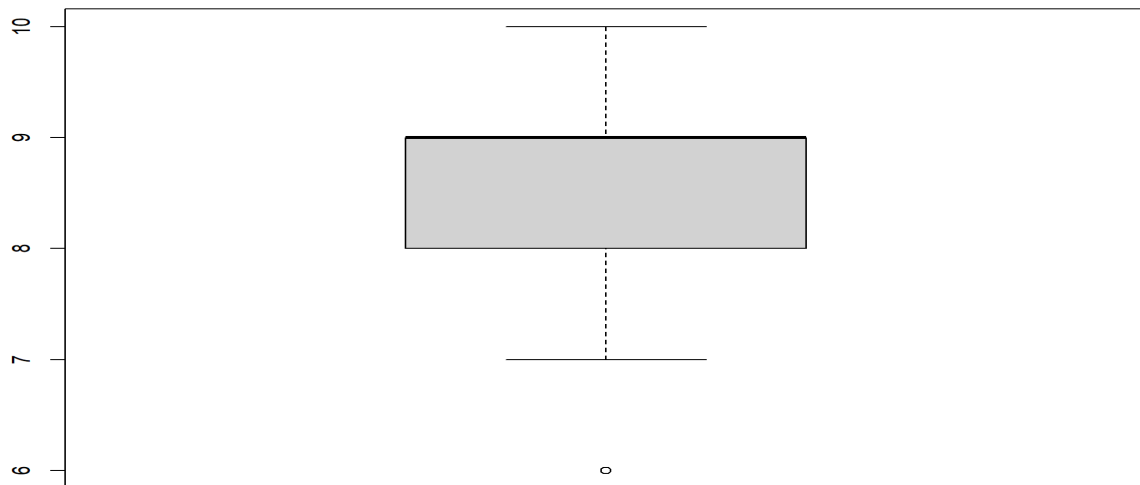
Boxplots are graphs that simply display the concentration of the data and highlight the specific outliers. It is of great importance to this analysis as it demonstrates the mean values and the dispersion of the data sets that were collected from the survey.

Boxplot of Age Beginning:



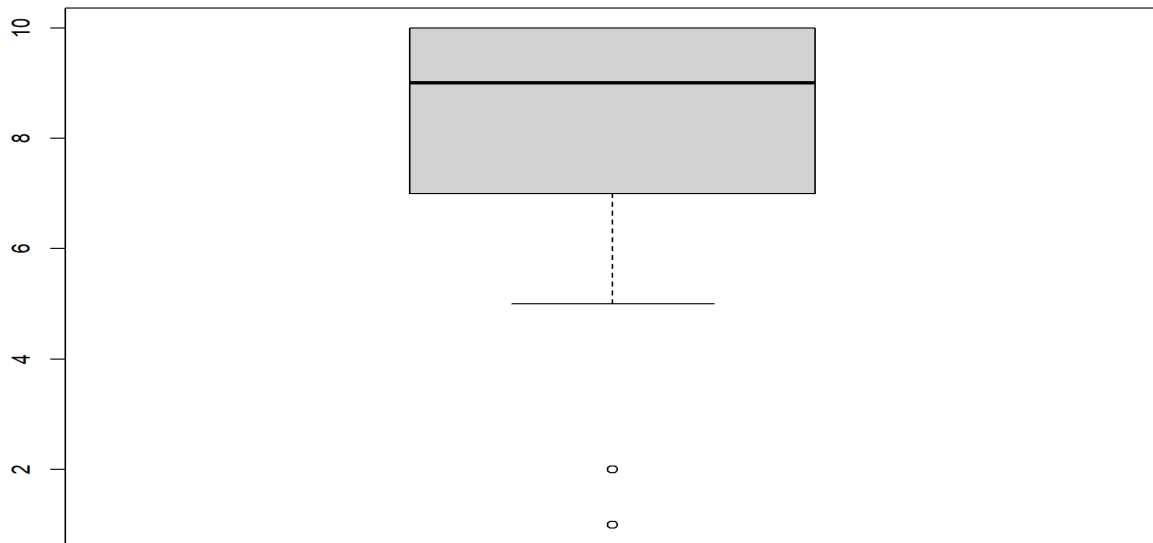
Graph 3 - Boxplot Age Beginning

The graph above shows the data dispersion and concentration under the variable Age Beginning. It is noticeable that, from the data set, there were three outliers above the upper tail. Meaning that there were three people who began their specialization significantly older than 25 years old. The rest of the answers were gathered around the mean, thus, showing that most people begin when they are in their early twenties. Once again, we can assume that the individuals that began later either had a change in career or moved to the Hospital Universitario de La Paz after completing their specializations at other hospitals.

Boxplot Importance of Size:**Graph 4 - *Boxplot Importance of Size***

This boxplot displays the data collected in regards to importance in size. As shown, the majority of the respondents centered their evaluation from 8 to 9. Above and below the central area are two symmetrical tails, meaning that it is neutrally skewed. Lastly, there is one outlier that rated a 6 for importance of hospital size. Even though there is one outlier, all of the data points are 6 and above, and most answers were focused on high ratings, meaning that importance in hospital size is a significant factor for students' choice for specialization.

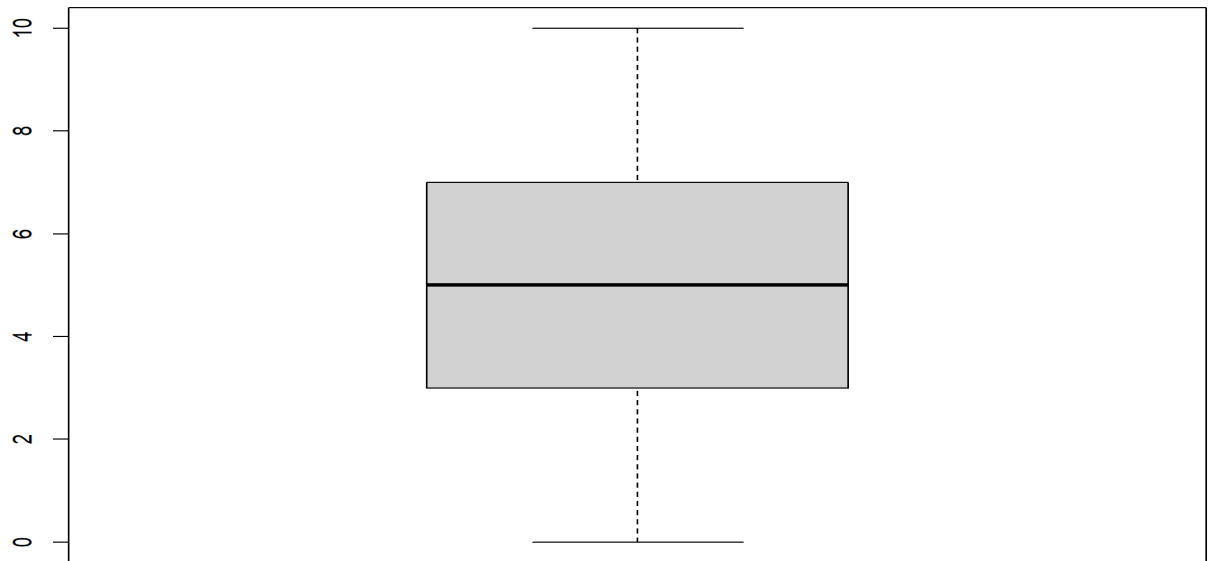
Boxplot Importance of Location:



Graph 5 - Boxplot Importance of Location

The boxplot above portrays the dispersion of the data collected in terms of importance of location. The box itself is concentrated around 7 through 10, with a tail extending to 5, and two major outliers. The intake we can absorb from this graph is that the location of the hospital is a major contributor to student's choice, even though there were two respondents who didn't seem to care much about this variable.

Boxplot Importance of Pay:



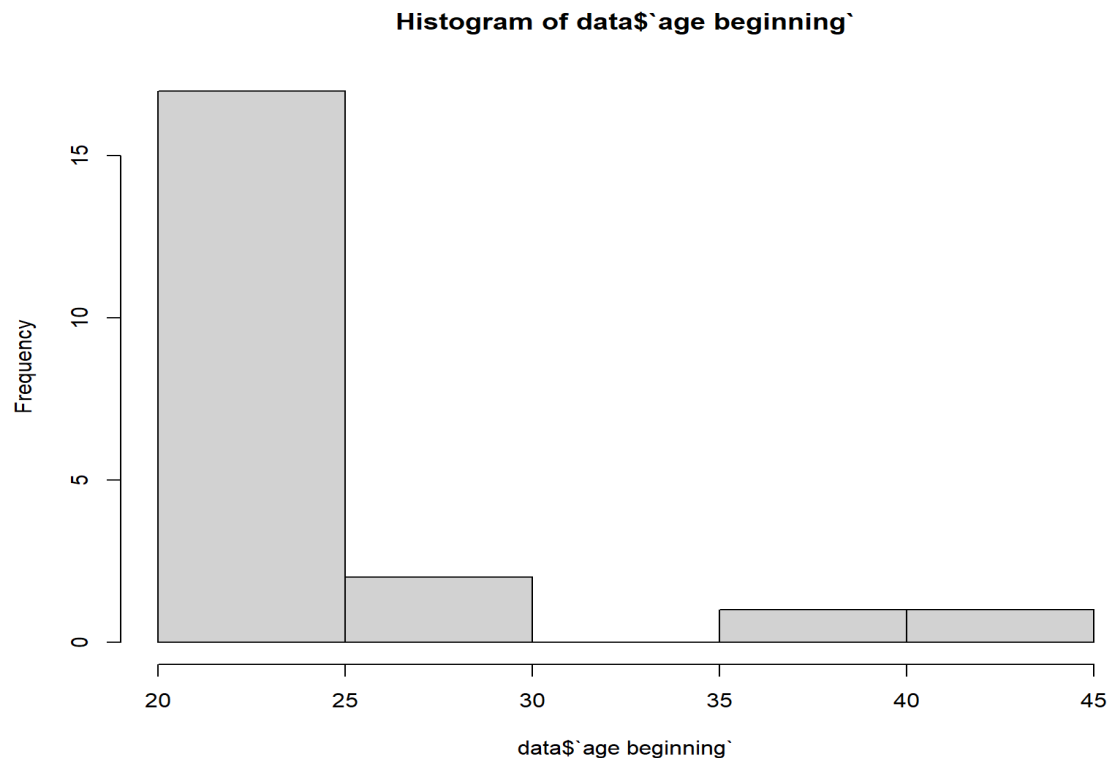
Graph 6 - Boxplot Importance of Pay

The boxplot above displays the data collected in terms of the variable importance of pay. It is noticeable that the data is centered between 3 and 7, with tails ranging from the whole data range. Given that, there are no outliers to be recorded. With this we can conclude that pay is not one of the main variables in the decision making process for these individuals as it ranges through 0 and 10, meaning, its importance changed for each individual and had no specific pattern.

Histograms of Data Collected

Histograms, in simple terms, are bar graphs that take into consideration one variable on the x-axis, and show the frequency of this given variable on the y-axis. This form of statistical analysis is useful for illustrating the distribution of the data in a simple way and to test for normality in the data.

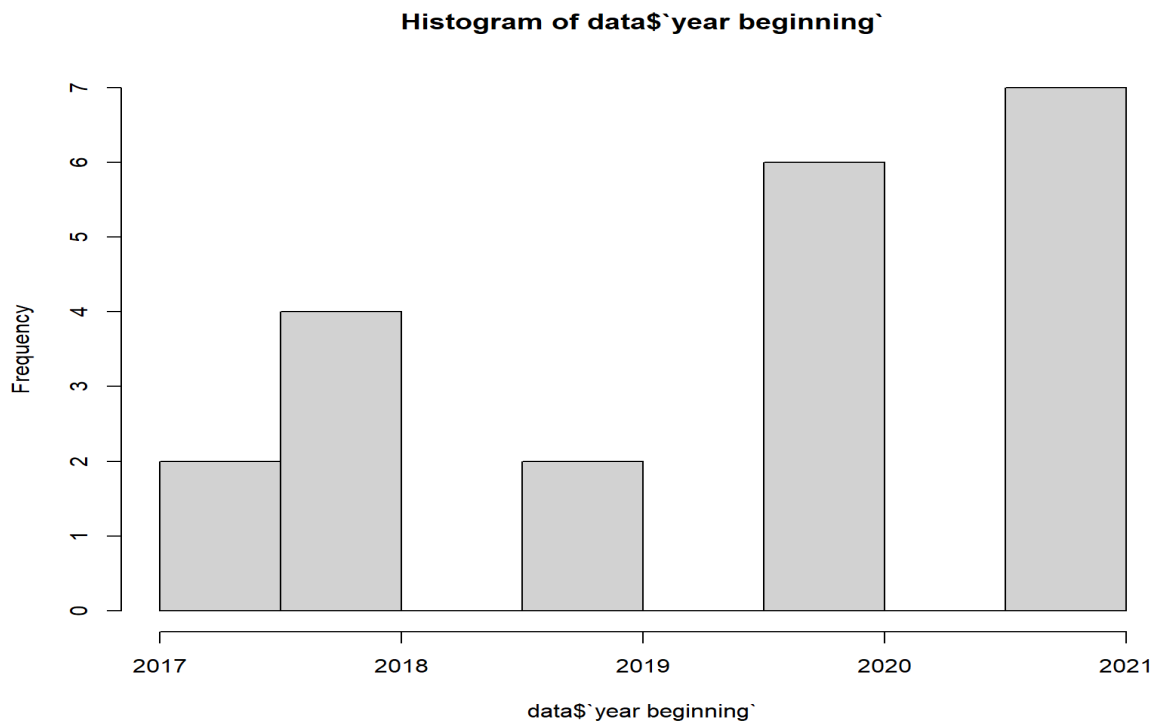
Histogram Age Beginning:



Graph 7 - Histogram Age Beginning

This given histogram displays the frequency of the age with which respondents began their specialization. From the graph, we can see that the range between 20-25 has a frequency exceeding 15, while the other ranges do not outgo the value of 3. This provides us the necessary information to conclude that most students begin to specialize in an area when they are between the ages of 20 to 25, thus, the hospital should target this age group for greater return. Furthermore, we can notice a strong right skew in the data, as pointed out by the quantity of individuals from each age group.

Histogram Year Beginning:

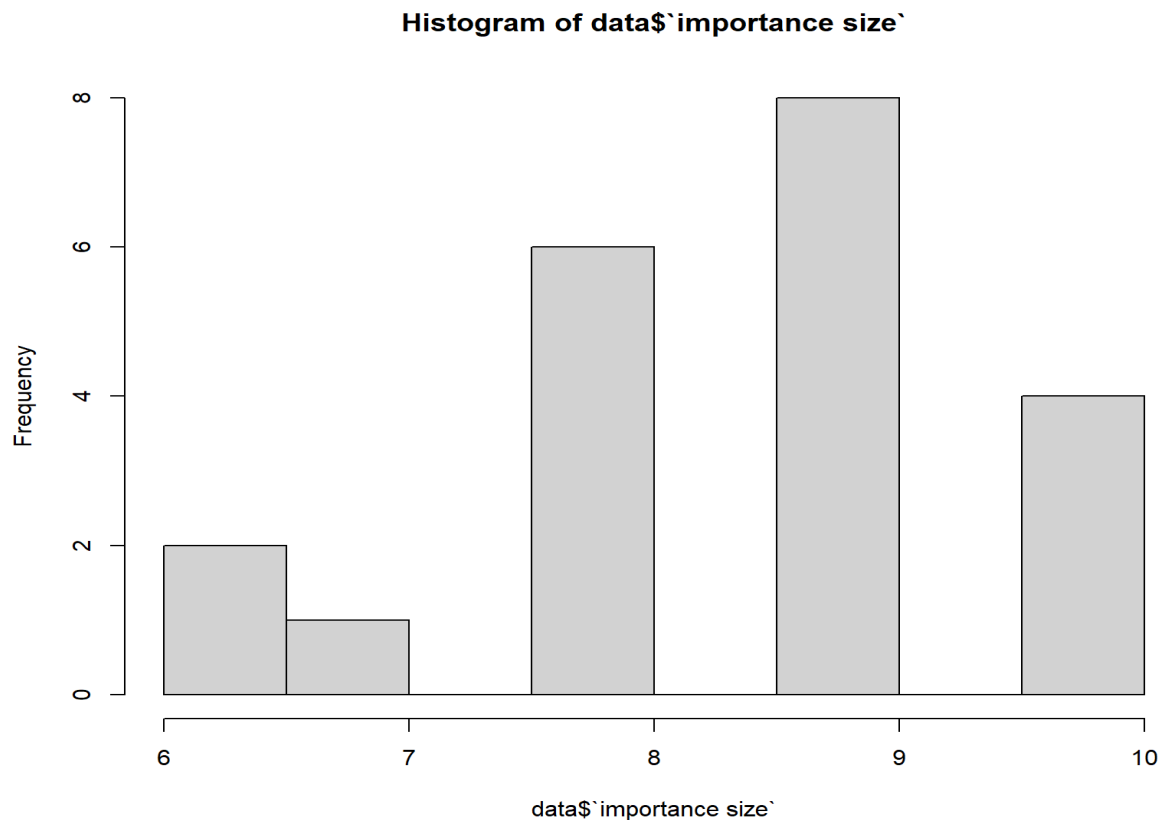


Graph 8 - Histogram Year Beginning

The histogram above depicts the frequency of the year students began their specialization. These variables cannot be significantly influenced by other factors, as the year students begin specialization is usually the year after they graduate from their previous courses. However, this graph does show useful information being that in the year of 2019, the frequency was very low, probably because of the COVID-19 pandemic. From the years past that, we can see that there's a considerable increase in the frequency. Therefore, we can acknowledge that the tendency is that the frequency will probably increase or keep stable the next following years as COVID becomes less prominent. Furthermore, once again, we can take into consideration that because most of the participants started in the years during the pandemic (2019-2021) the conclusions from this report may not apply to others starting before this pandemic. On the other hand, as the world recovers from this virus, many of its repercussions will remain. Because of this, we can predict that the changes in important

factors seen in the years during the pandemic may remain as one of the most important although it might decrease a small percentage.

Histogram Importance Size:



Graph 9 - Histogram Importance Size

This histogram shows the frequency of the importance of the size variable. As shown, it has a weak left skew where the highest point is at 9. This demonstrates that importance in size is also a major determinant in the matter.

Correlation of Data using Correlation Matrix

The Correlation Matrix is a data analysis technique where all of the possible variable combinations are evaluated, measuring the relationship between them. Understanding this

correlation is fundamental to our analysis since we need to know what are the main determinants of students' hospital choices. This is shown in the figure below.



Graph 10 - Correlation Matrix of all Variables

Above is displayed the Correlation Matrix graph. To better understand this graph, it's important first to recognize that the Y-axis and the X-axis display the variables used, and each box shows how strong the relationship between them is. On the right of the graph, we provided a tool explaining the colors assigned to each strength level, being 1 completely correlated and 0 not correlated at all.

Given this graph, a few points stand out that are relevant to our analysis. The first one is the relationship between the importance of location and the importance of pay, where there's a moderately high correlation (that of 0.5576). This reveals that one has a significant effect on the other, having a p-value of 0.009, meaning that students are likely to say pay is an important factor when location is and vice-versa. However, the hospital payments in Spain are standardized from hospital to hospital, only changing from specialization type. Thus, the

relevant piece of information that we can extract from this graph and use is that salary tends to have smaller importance when the hospital is well located. More information about the location is given in the results section. A second important point is that the importance of location and the year the students begin their specialization are also moderately highly correlated (with a correlation of 0.5748) and significant with a p-value of 0.006. Possibly in the last few years location has become more important as the COVID pandemic has made it so individuals might not want to go very far from their homes or want to be with their families or in these main cities where it was easier to get around .

Thirdly, it is important to understand that multiple other variables do have a weak correlation with one another (that is less than 0.40) meaning they have a very weak relationship or none at all. Because of this, there should be no interaction or impact of one variable on another. They may be important in the final decision of the hospital workers when choosing a hospital but we do not expect to see any relationships among these variables in the graphical analysis completed below.

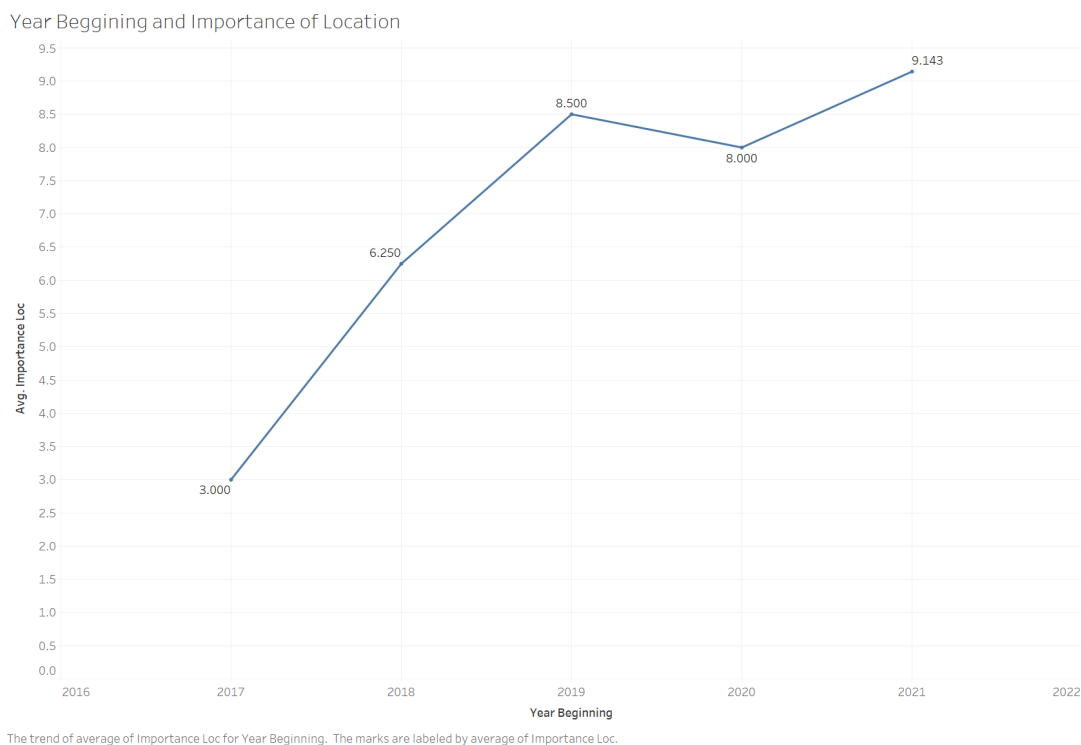
Further Analysis

After recollecting all the answers of the surveys and putting them into an excel, we were able to analyze the distinct factors individuals take into consideration when deciding the hospital where completing their specialization at. Some of the patterns that can be seen based on the data is that most of the people believe the most important factors when deciding where to do their specializations are based on the location of the hospital, good working environment, and its prestige. Due to this, they chose Hospital Universitario La Paz because it was the one that met most of the factors they were considering. Furthermore, when deciding their specializations the factors they based on were very similar to the ones stated above. The most important factors they considered for this were their personal interest in the

specialization itself, the job opportunities, and the quality of life. On the other hand, in the way they gained information about the hospital's training quality there was a similar pattern between all the participants of the survey since most of them responded that they gained this information from former hospital staff and residents. Moreover, their answers to the survey included this factor with other sources of information which vary among the individual responses. As a result, most of the answers of the 21 participants of the surveys were very similar despite them having different specializations.

Taking into consideration that location was one of the key factors in the decisions of hospital workers, we decided to dive deeper into how this takes form and plays into them choosing Hospital Universitario de La Paz concretely. The average distance from Hospital Universitario de La Paz is higher than the mean distance of hospitals (24 minutes) from the center of Madrid , Sol, with a time of 27 minutes by metro. The minimum is 13 minutes and the maximum is 39 minutes so there seems to be a slight skew in the distribution of the times, although small. Due to this, we can assume that individuals put location as one of the more important factors for their decision in choosing this hospital based on the fact that Madrid is a large city and the capital of Spain. Both in the “3 most important factors” and “how important is location” questions we found that the location of the hospital was a key factor, having a mean importance level of 9/10. Since individuals tend to prefer shorter distances, we recommend that in the future if there were to be construction of more hospital buildings, putting them closer to the center (Sol) or near current buildings and make sure that it is still well connected by metro. We did see that some hospitals required multiple sources of commute if not gone by car or walking including the renfe cercanias, buses and metros. This could possibly decrease the motivation of individuals, so we can infer that the connection between the center and the Hospital Universitario de La Paz, only needing metro, is more comfortable and an attractive option. On the other hand, Nicole Moller (2021) and others

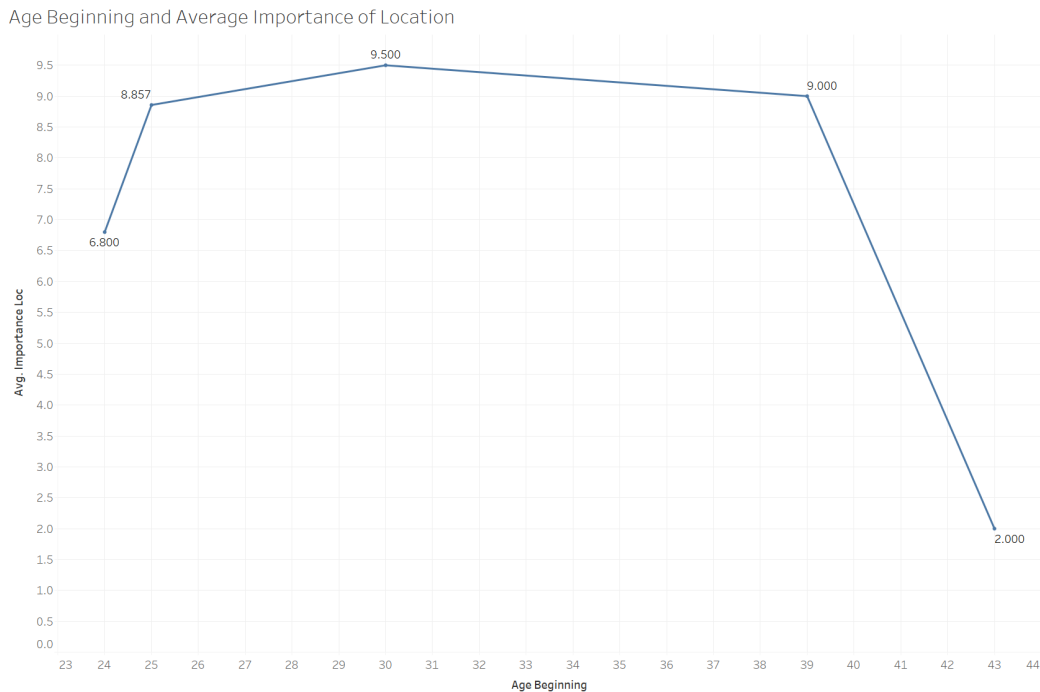
spoke about the fact that nurses are willing to commute to larger cities and capital cities because of the possibly better opportunities for themselves and unique environments in the hospitals in the city. Inside Madrid itself, we can infer that individuals choose Hospital Universitario de La Paz for the same reasons; these being the level of opportunities and other unique features, including quality of training, quality of life, etc, regardless of having a longer distance from the city center.



Graph 11 - Graph of Year Beginning Specialization and Importance of Location

When analyzing the line graphs about year beginning vs importance of location, shown above, it can be seen that there is a correlation between them since both of them tend to increase when the other does. This might be due to the fact that in the last years location has become one of the most important factors to take into consideration since Madrid is a very large city with lots of citizens. Still though, it can be seen in the graph that despite the overall increase through the years, there was a decrease in both year beginning and in the average of location importance in 2020, which might be due to the pandemic. Despite this, in

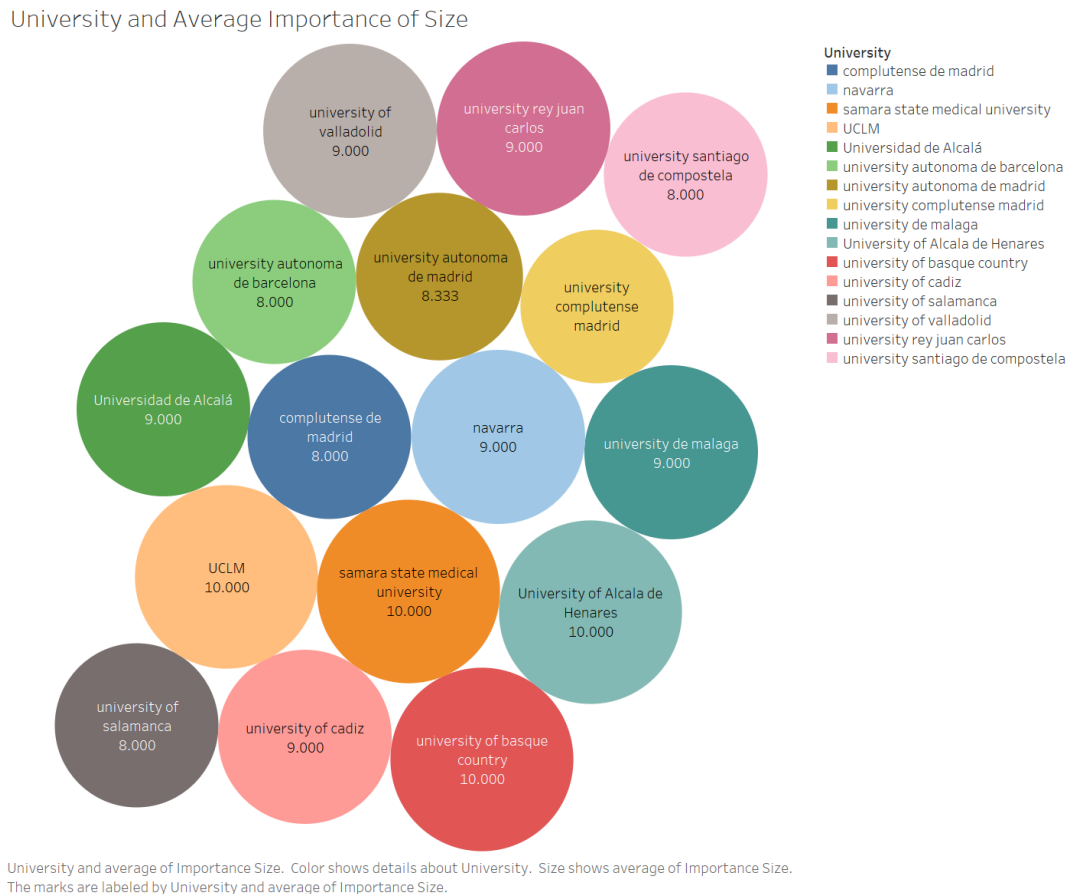
the following year, there was a significant increase in both of these factors which has been the highest so far.



The trend of average of Importance Loc for Age Beginning. The marks are labeled by average of Importance Loc.

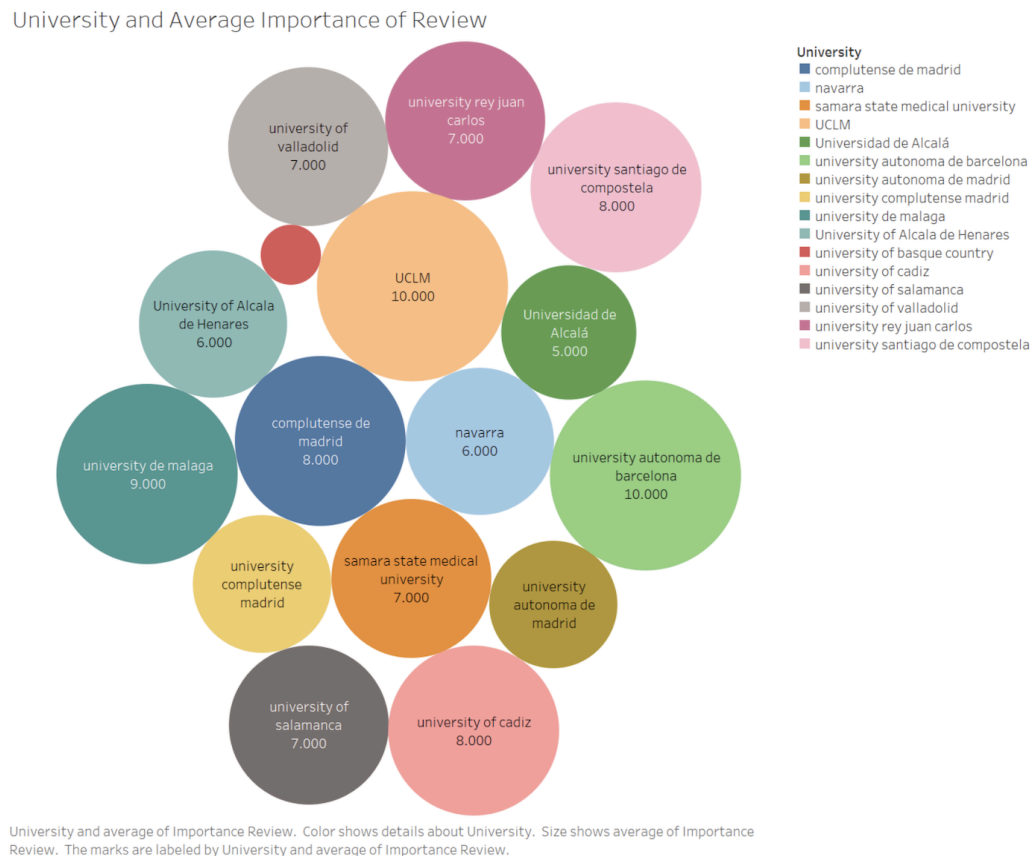
Graph 12 - Graph of Age Beginning Specialization and Importance of Location

On the other hand, when looking at the graph above which compares the age beginning vs the average importance of location, there isn't an identifiable pattern. At the beginning of the graph it can be seen that there is a correlated increase between the two factors analyzed but then from the age beginning of 30 there starts to be a decrease of the average importance of location as the age of beginning is higher. This might be due to the fact that as older is the age of beginning, the importance of location tends to be lower than younger ages of beginning.



Graph 13 - Graph of University Graduated from and Importance of Size

Furthermore, when analyzing the university versus the average importance of size, it can be portrayed that the university students that care the most about the size are from universities in Spain that are not located in Madrid. This might be because as they come from other cities in Spain, these individuals tend to give more importance to larger hospitals when considering the ones that are not located in their hometown. Furthermore, we can suspect that due to the size of Madrid and it being the capital city, many individuals may be drawn to work in such a space. Still though, the graphs show that universities in Madrid have similar average importance of size but still are lower than the ones of other universities in Spain which are not located in Madrid.



Graph 14 - Graph of University Graduated from and Importance of Review

Moreover, the graph below which is similar to the one analyzed above compares the universities versus the average importance of review. It can be seen that the outcomes are similar since most of the universities which have the highest average importance of review are the ones which are not located in Madrid, but in other parts of Spain. Still though, there are some universities such as Complutense de Madrid which also have a high average importance of review and higher than other universities that are not located in Madrid. This might be due to the fact that individuals take into consideration all kinds of factors such as location. Additionally, it can be speculated that those in Madrid, are already surrounded by high level hospitals so their concern for the importance of review is not as high as they may feel that wherever they will go, the quality will be good.



Graph 15 - Graph of University Graduated from and Importance of Location

Finally, when comparing the university vs the average importance of location, we were able to prove that this is one of the most important factors. As seen in the graph, the universities which have the highest average importance of location reach a 10.0 in the scale and are not located in Madrid. So, in the deciding process of hospitals, for one to choose a hospital in a different city it will need to have a good location, once again relating to the idea of the metro system and distances inside the city of Madrid. This factor is of huge importance for individuals because it's what motivates them to choose one hospital over another. Also, it must remain in consideration that Madrid being one of the biggest cities in Spain can also influence the decision process when choosing between cities and hospitals where to work.

Conclusion

After researching and analyzing the data, we were able to identify and meet our objective of finding the main determinants at the student and hospital level of hospital choice for doctor's specialization in Madrid, focusing mainly on Hospital Universitario de La Paz residents. When taking into consideration several literature reviews from current hospitals in Spain, it was seen that the most common variables for evaluating training quality was case mix and workload. Despite this, we wanted to discover the main determinants of choice for candidates of HULP by conducting a survey to its residents. The results of the mentioned survey weren't as favorable as we would have expected since only 21 individuals responded to it out of around 350 residents. Due to this, we did not have enough data to conduct a proper analysis of the factors that medical students take into consideration when deciding their specialization and hospital to complete these at. Still, with the data we were able to collect, it can be seen that the 3 most common factors when choosing a hospital to complete the specialization at were location, good work environment, and reputation and prestige. Furthermore, the 3 most important factors when deciding on specializations were the interest in the speciality itself, the job opportunities, and the quality of life.

As a recommendation for future analysis, it would be best if more individuals from Hospital Universitario de La Paz responded to the survey since this will allow us to conduct a deeper analysis of the affecting factors when making the decision of choosing a hospital over another. Additionally, Hospital Universitario de La Paz should focus on spreading the information about their hospital through other sources such as social media, because most of the respondents from the survey stated that they gained the information about the hospital through hospital staff and residents or previous students more than they did from media. Moreover, the hospital should try to focus more on engaging medical students from Madrid to do their specializations in Hospital Universitario de La Paz since the graphs mentioned before

show that most of the respondents of the survey are from other cities in Spain. In addition to this, there would be a positive correlation with Hospital Universitario de La Paz adding more buildings in Madrid increasing their level of choice over other hospitals. This can be due to the distinct locations in such a big city as Madrid, benefiting individuals living farther away leading to a smaller commute. Since location is one of the most common factors when choosing a hospital to complete specializations at, this implementation would have a huge impact on the hospital itself.

References

- Aagaard, E. M., Julian, K., Dedier, J., Soloman, I., Tillisch, J., & Perez-Stable, E. J. (2005). Factors affecting medical students' selection of an internal medicine residency program. *Journal of the National Medical Association*, 97(9), 1264.
- Agawu, A., Fahl, C., Alexis, D., Diaz, T., Harris, D., Harris, M. C., ... & Higginbotham, E. J. (2019). The influence of gender and underrepresented minority status on medical student ranking of residency programs. *Journal of the National Medical Association*, 111(6), 665-673.
- Freire, J. (2015, June 2). *An analysis of the medical specialty training system in Spain - Human Resources for Health*. SpringerLink. Retrieved April 12, 2022, from https://link.springer.com/article/10.1186/s12960-015-0038-y?error=cookies_not_supported&code=813754af-baf0-4175-8205-1988452ad52a
- Harris, J. E., Lopez-Valcarcel, B. G., Barber, P., & Ortun, V. (2016, February 16). *Allocation of Residency Training Positions in Spain: Contextual Effects on Specialty Preferences*. NIH- National Library of Medicine. Retrieved April 12, 2022, from <https://pubmed.ncbi.nlm.nih.gov/26880315/>
- Kazerooni, EA; Blane, CE; Schlesinger, AE; Vydarney, KH. Medical students' attitudes toward radiology; comparison of matriculating and graduating student. *Academic Radiology*. 8/1/1997. 4(8). 601-607.
- Machado, M. P., Mora, R., & Romero-Medina, A. (2012). Can We Infer Hospital Quality from Medical Graduates' Residency Choices?. *Journal of the European Economic Association*, 10(6), 1400-1424.
- Moller, N., Berthelsen, C., & Holge-Hazelton, B. (2021, December 30). Driving for the unique opportunity for work: a qualitative study of nurses' motivation to commute to work | Emerald Insight. Emerald Insight. Retrieved June 8, 2022, from

https://www.emerald.com/insight/content/doi/10.1108/JMP-02-2021-0107/full/html?casa_token=Gx1c4fk8ruMAAAAAA:rOCkesOdKIhH5vI2-VVu73SZHfBGPIFaksXpudsoXYdlGVGHTTWrNqcDY3LuMZ4o3UbGHFNOa-q-7Yib5IiahXF5y00F6f-OiDSSAPYoBn7TiSRO3eB8

Wang, T., Wong, B., Huang, A., Khatri, P., Ng, C., Forgie, M., ... & O'Neill, P. J. (2011).

Factors affecting residency rank-listing: a Maxdiff survey of graduating Canadian medical students. *BMC Medical Education*, 11(1), 1-7.

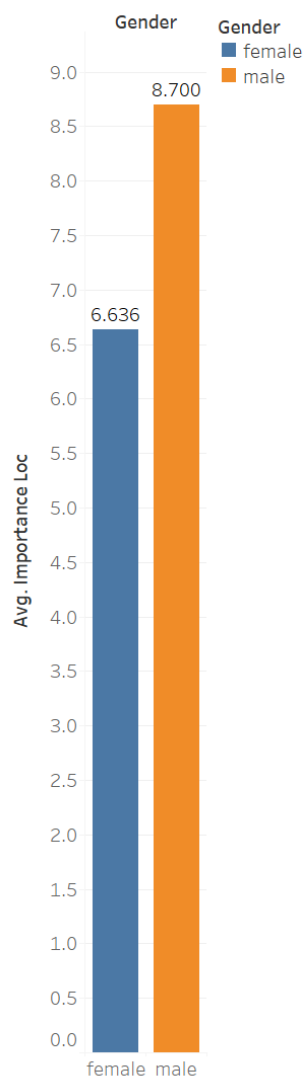
Appendix

Link of Survey for Hospital Universitario de La Paz Staff:

<https://forms.gle/AbrgpCT7VtLCV8cM6>

Link of Survey for other hospitals' staff: <https://forms.gle/2kiDXeXDVbJMmwR99>

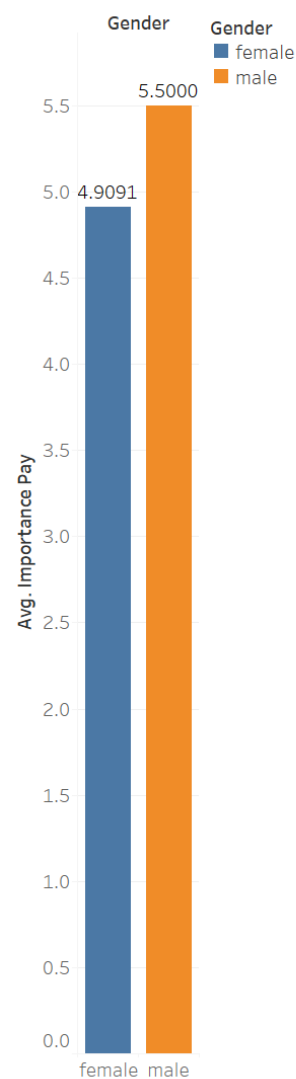
Gender and
Average Im-
portance of
Location



Average of Importance
Loc for each Gender.
Color shows details
about Gender. The
marks are labeled by
average of Importance
Loc.

Graph 16 - Graph of gender and importance of location

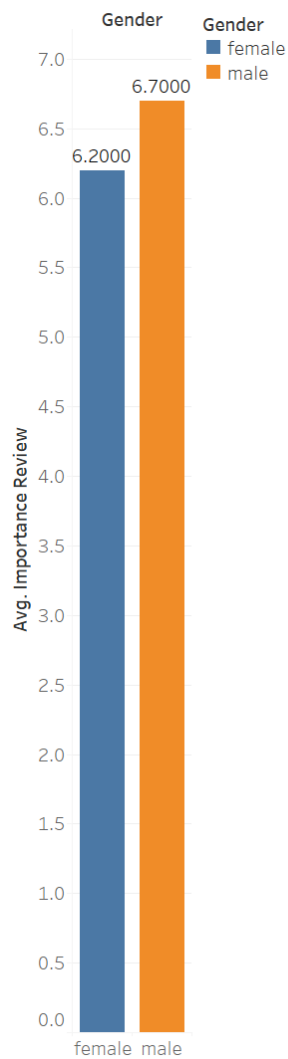
Gender and
Average Im-
portance of
Pay



Average of
Importance Pay for
each Gender. Color
shows details about
Gender. The marks
are labeled by average
of Importance Pay.

Graph 17 - Graph of gender and importance of pay

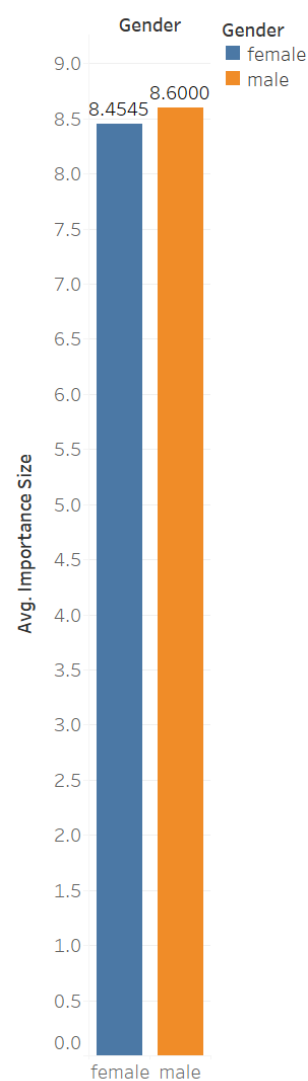
Gender and
Average Im-
portance of
Review



Average of
Importance Review
for each Gender.
Color shows details
about Gender. The
marks are labeled by
average of Importance
Review.

Graph 18 - *Graph of gender and importance of review*

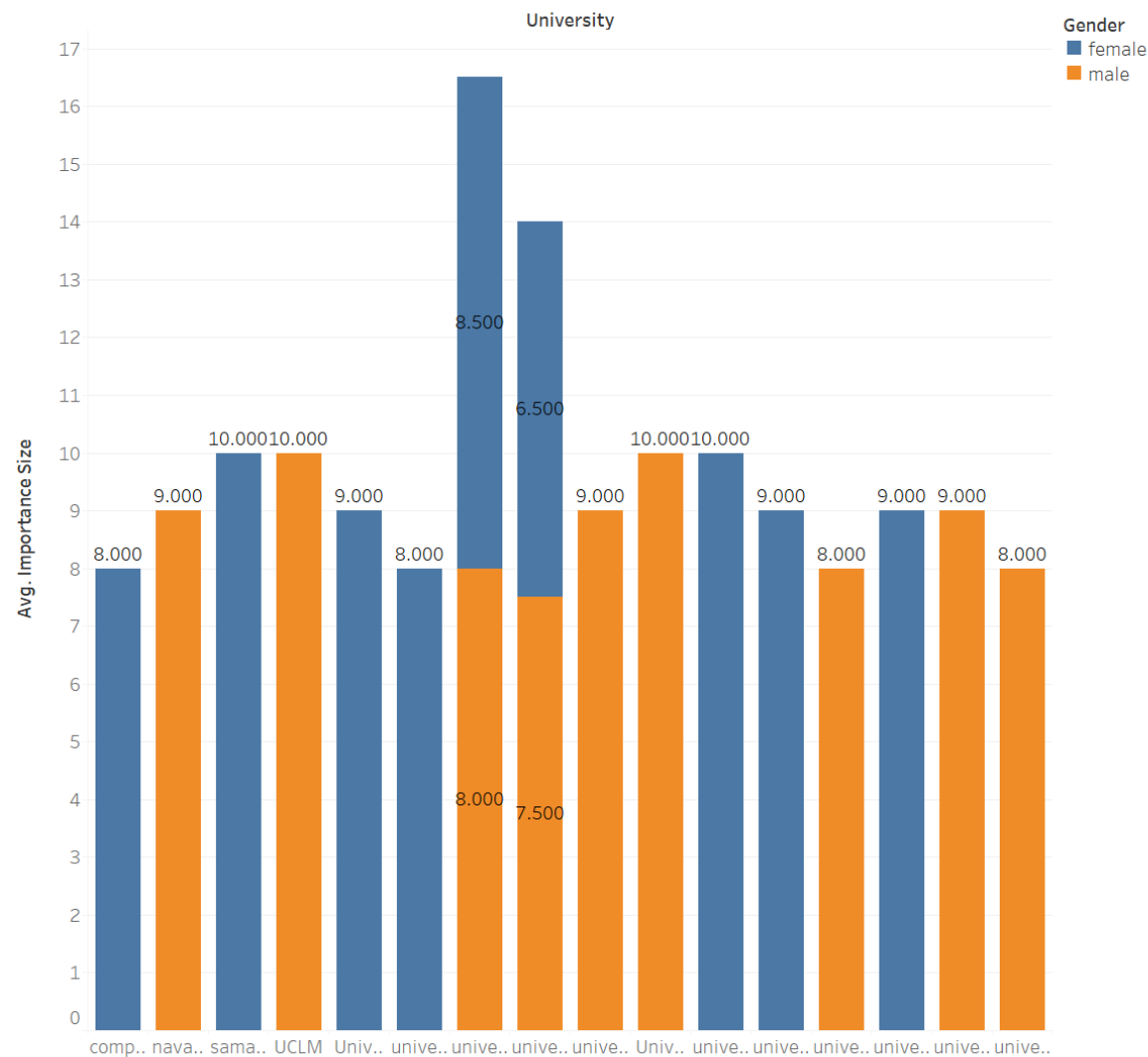
Gender and
Average Im-
portance of
Size



Average of Importance
Size for each Gender.
Color shows details
about Gender. The
marks are labeled by
average of Importance
Size.

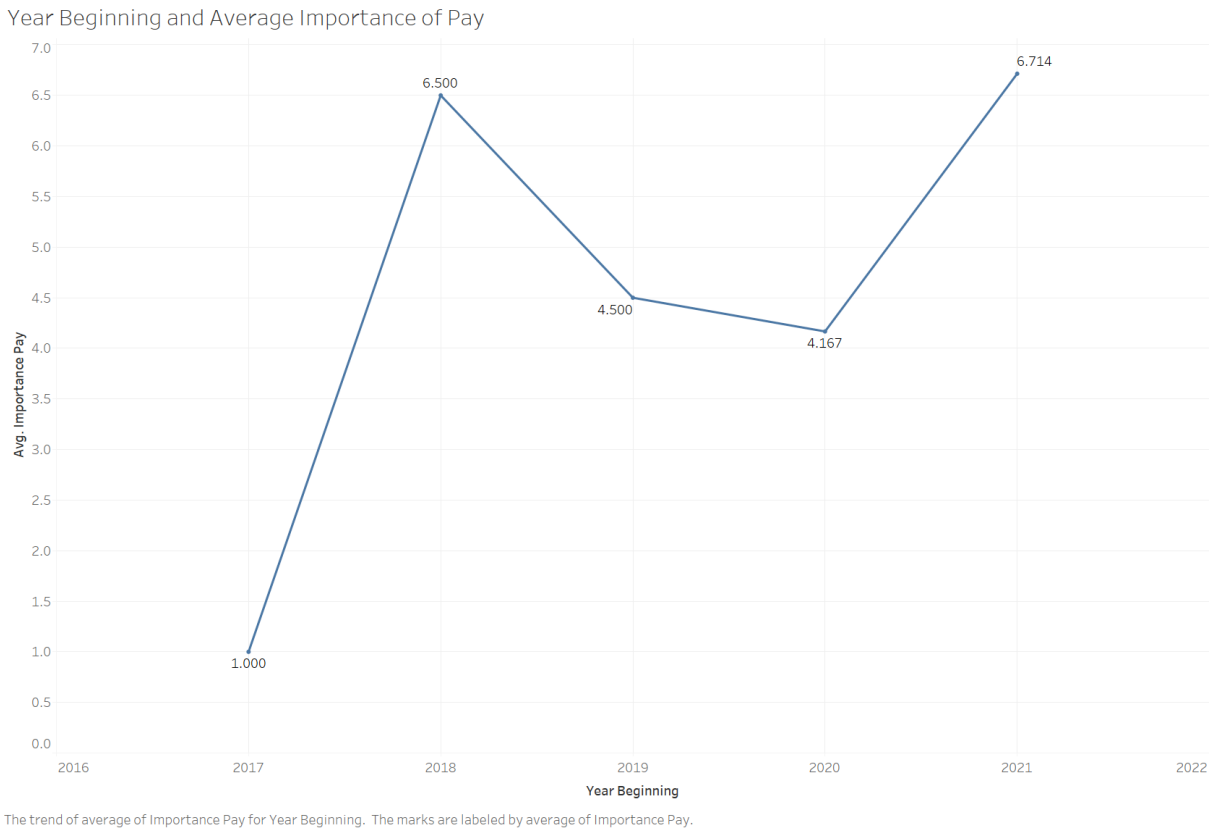
Graph 19 - Graph of gender and importance of size

University and Average Importance of Size Based on Gender

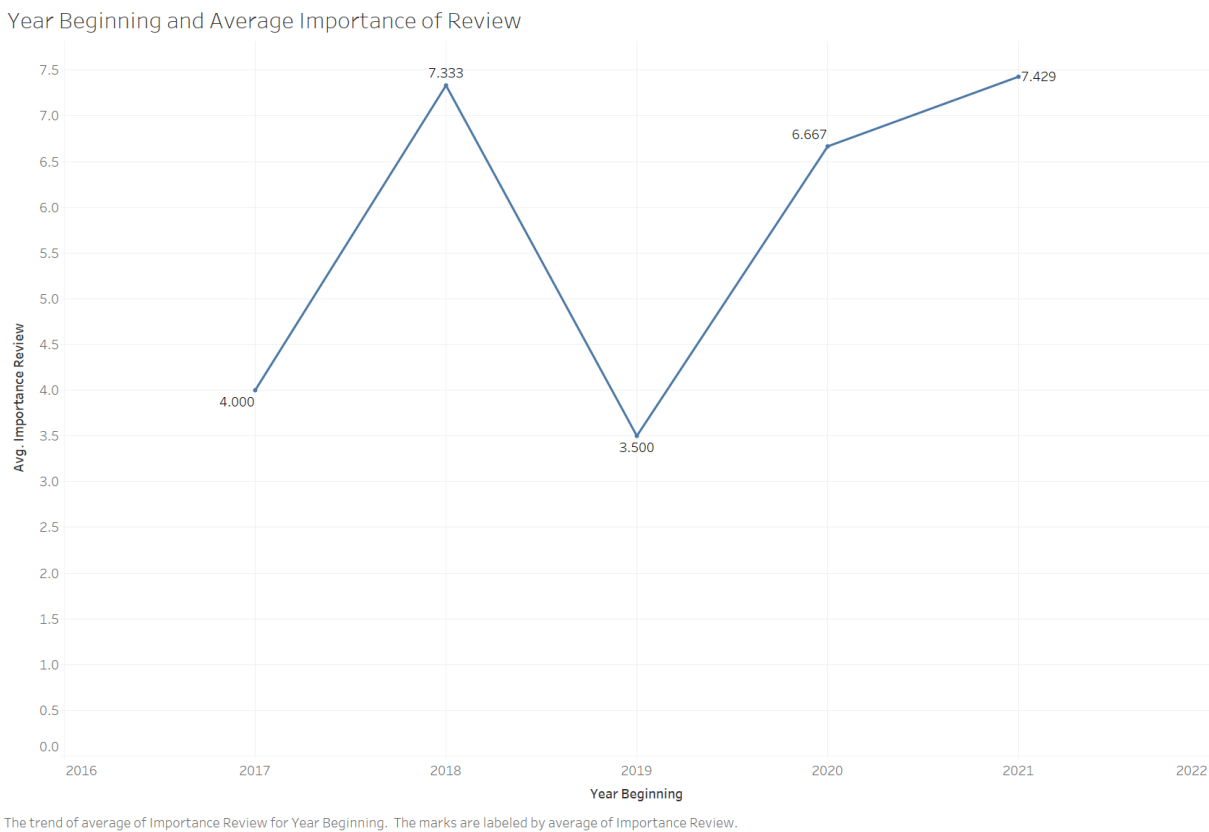


Average of Importance Size for each University. Color shows details about Gender. The marks are labeled by average of Importance Size.

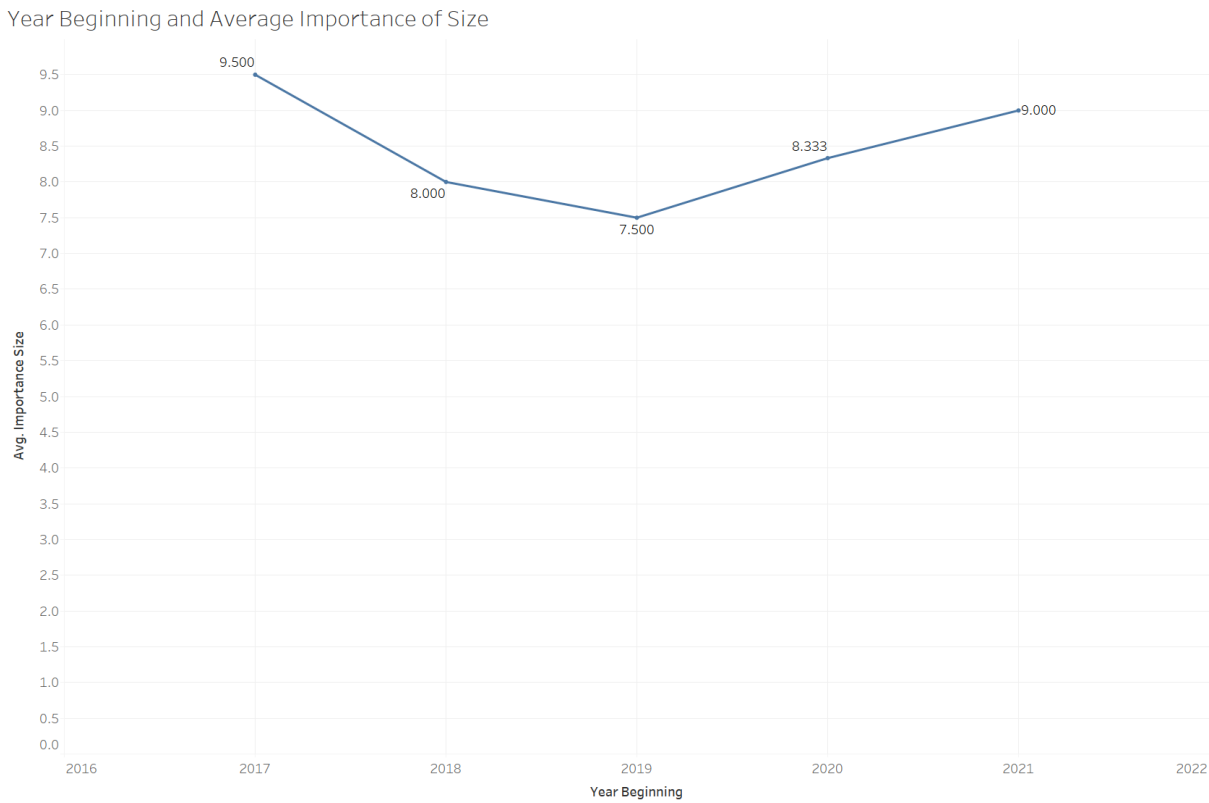
Graph 20 - Graph of gender and importance of size by university



Graph 21 - Graph of year beginning and importance of pay



Graph 22 - *Graph of year beginning and importance of review*



The trend of average of Importance Size for Year Beginning. The marks are labeled by average of Importance Size.

Graph 23 - *Graph of year beginning and importance of size*

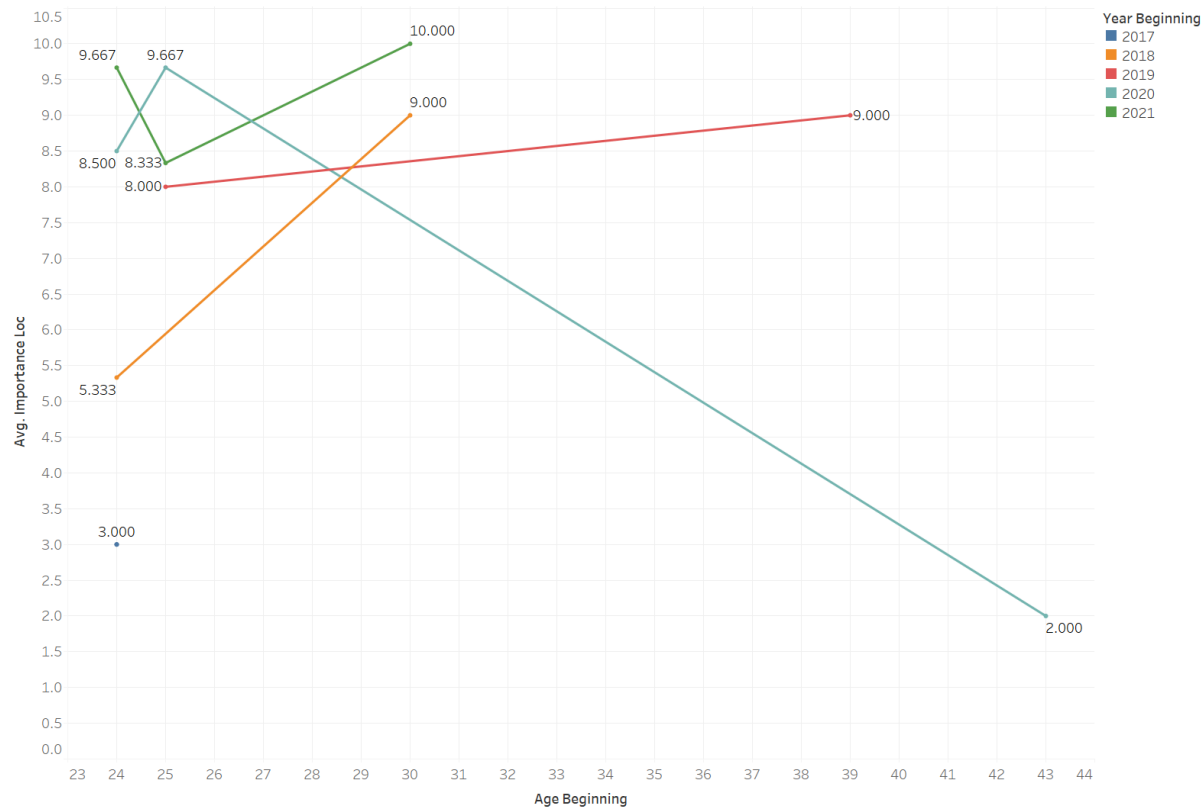
University and Average Importance of Pay



University and average of Importance Pay. Color shows average of Importance Pay. Size shows average of Importance Pay. The marks are labeled by University and average of Importance Pay.

Graph 24 - Graph of University Graduated from and Importance of pay

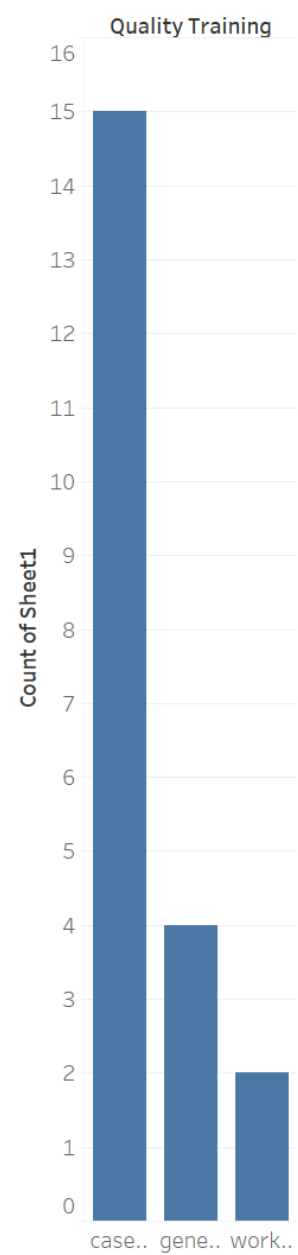
Age Beginning and Average Importance of Location based on Year Beginning



The trend of average of Importance Loc for Age Beginning. Color shows details about Year Beginning. The marks are labeled by average of Importance Loc.

Graph 25 - Graph of age beginning and importance of location depending on year beginning

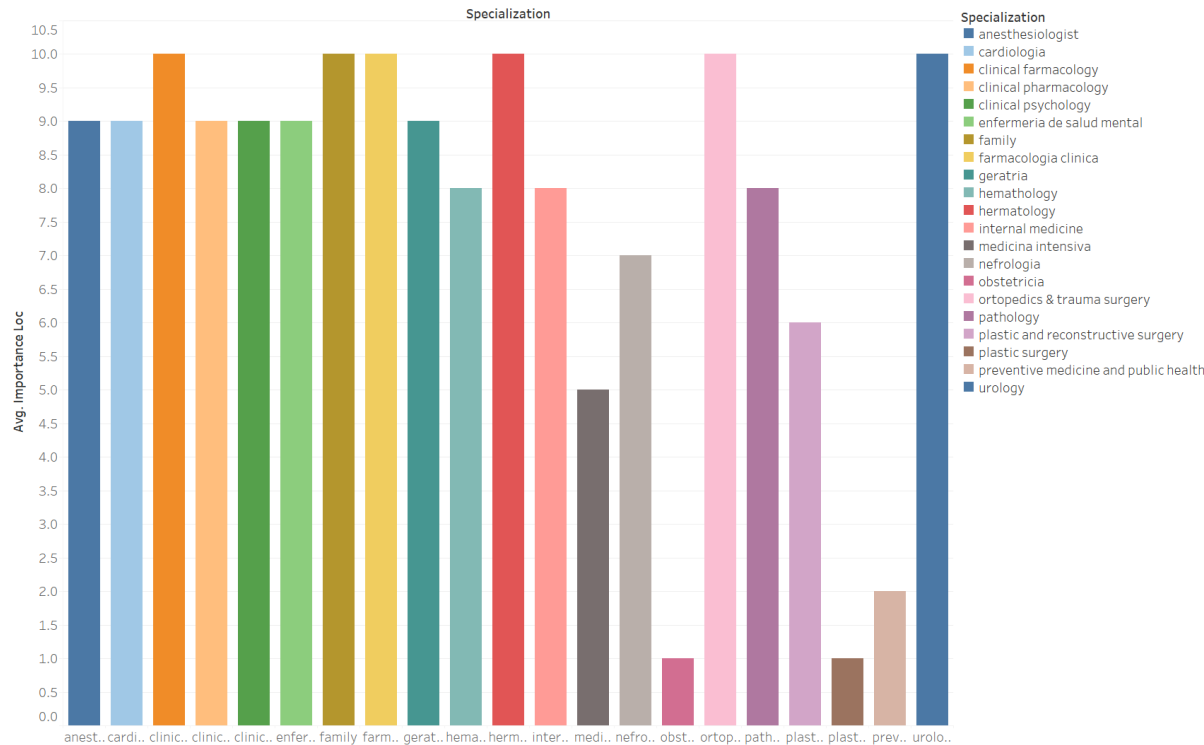
Definition of
Quality Training



Count of Sheet1 for each
Quality Training.

Graph 26 - *Graph of definition of training quality*

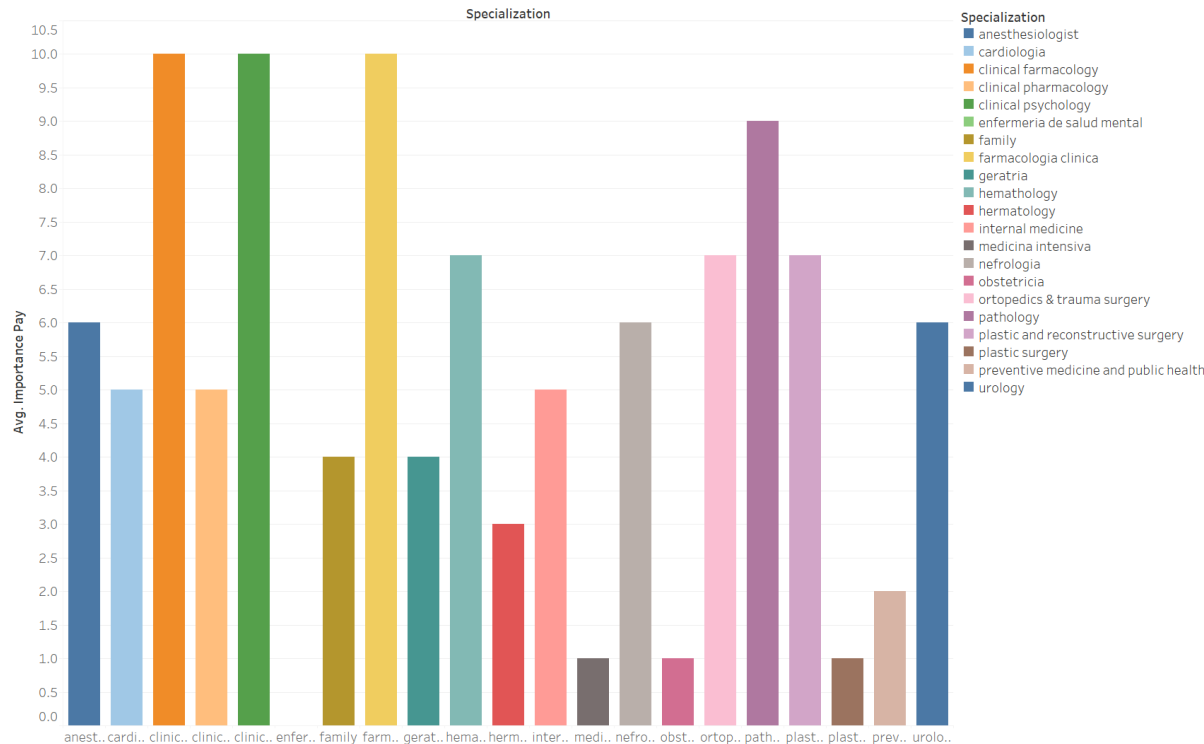
Specialization and Average Importance of Location



Average of Importance Loc for each Specialization. Color shows details about Specialization.

Graph 27 - Graph of specialization and Importance of Location

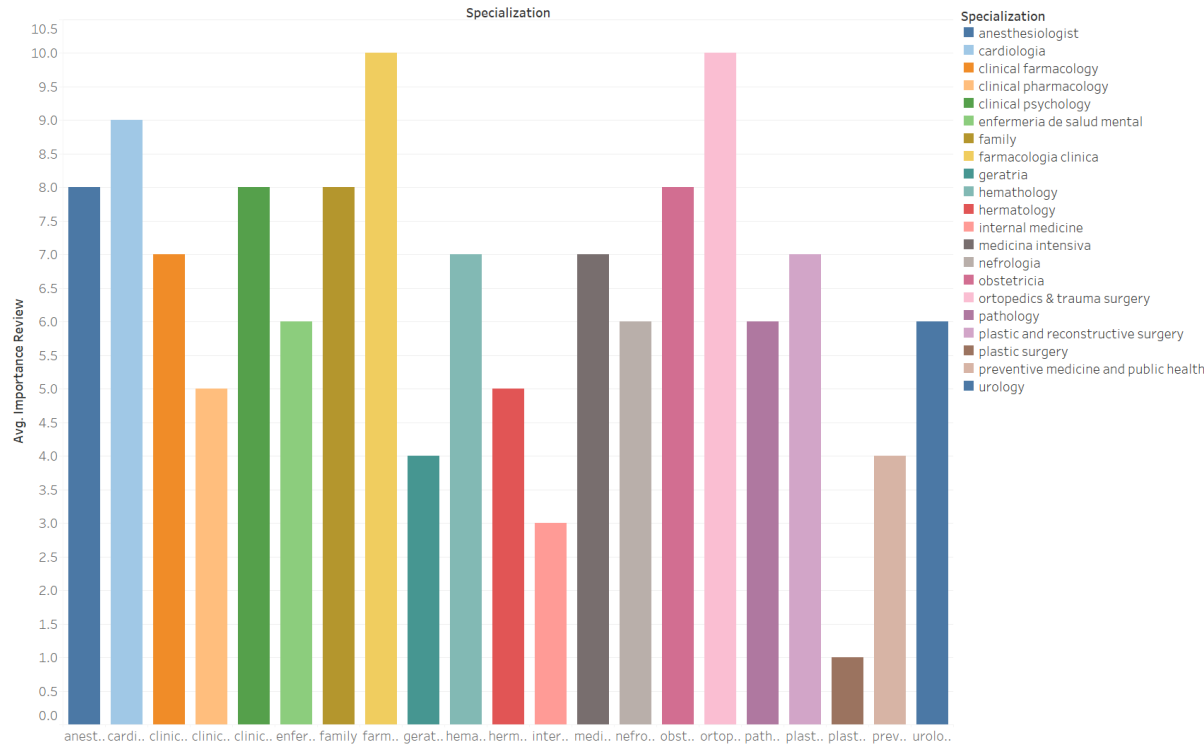
Specialization and Average Importance of Pay



Average of Importance Pay for each Specialization. Color shows details about Specialization.

Graph 28 - Graph of specialization and Importance of pay

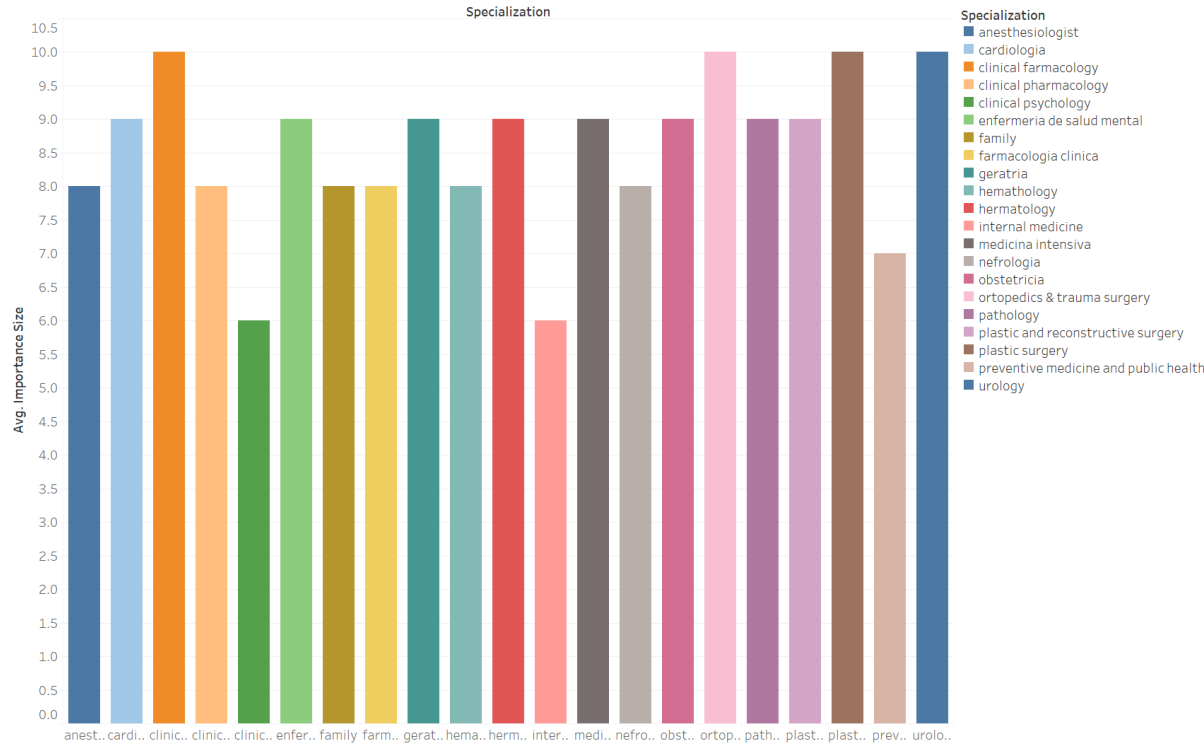
Specialization and Average Importance of Review



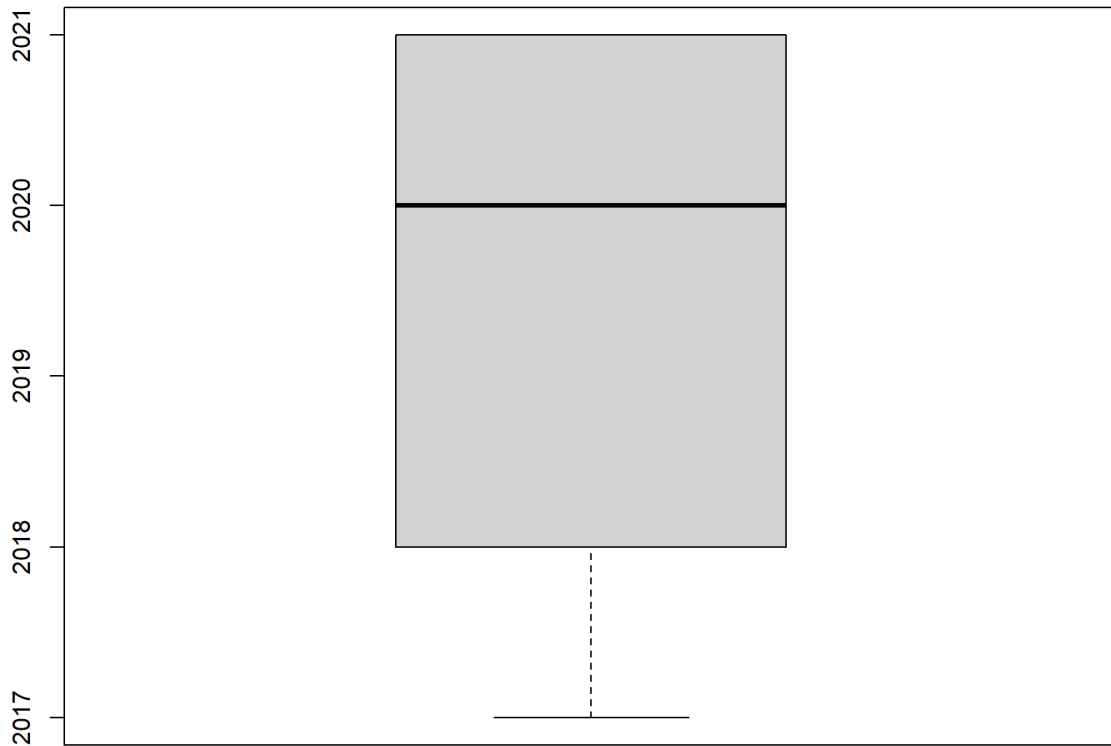
Average of Importance Review for each Specialization. Color shows details about Specialization.

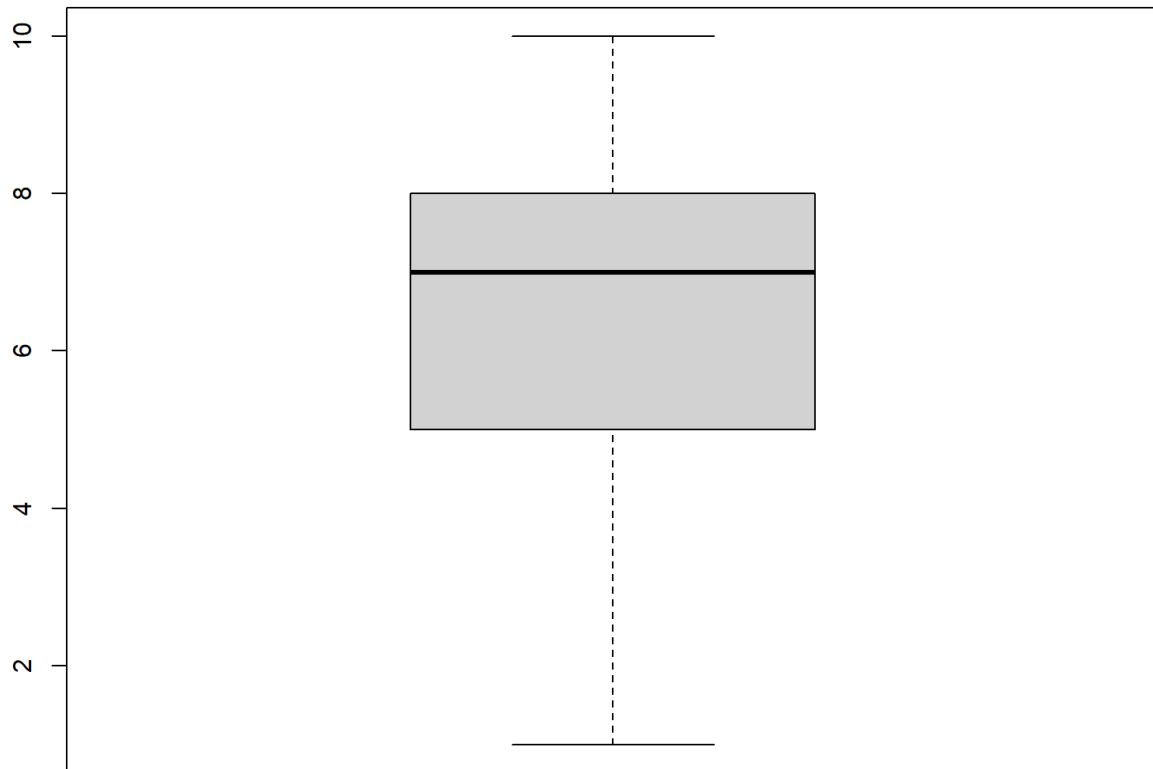
Graph 29 - Graph of specialization and Importance of review

Specialization and Average Importance of Size

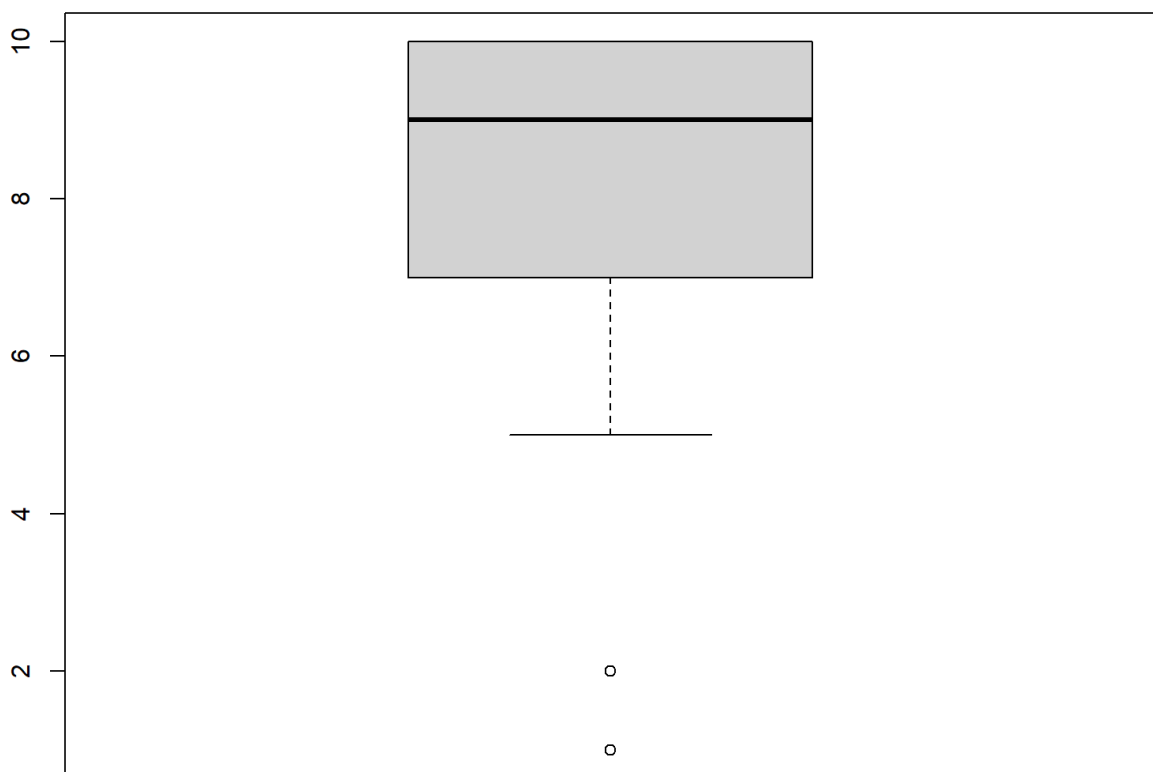


Average of Importance Size for each Specialization. Color shows details about Specialization.

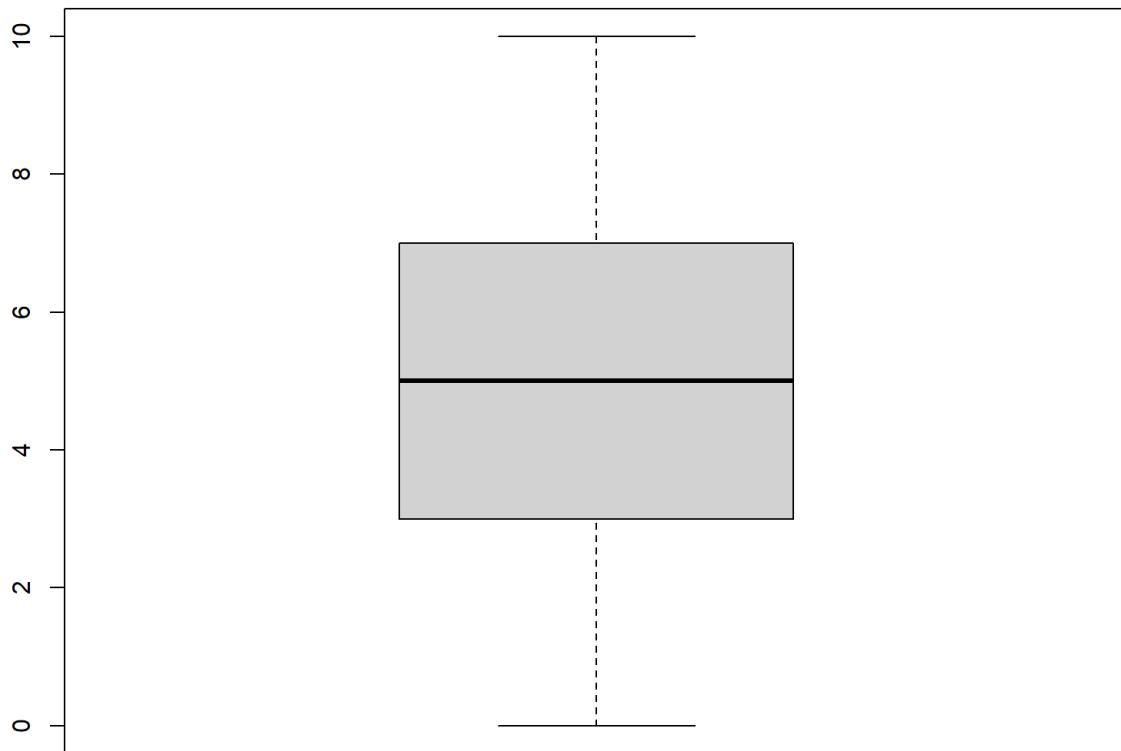
Graph 30 - *Graph of specialization and Importance of size***Graph 31** - *boxplot of Importance of size*



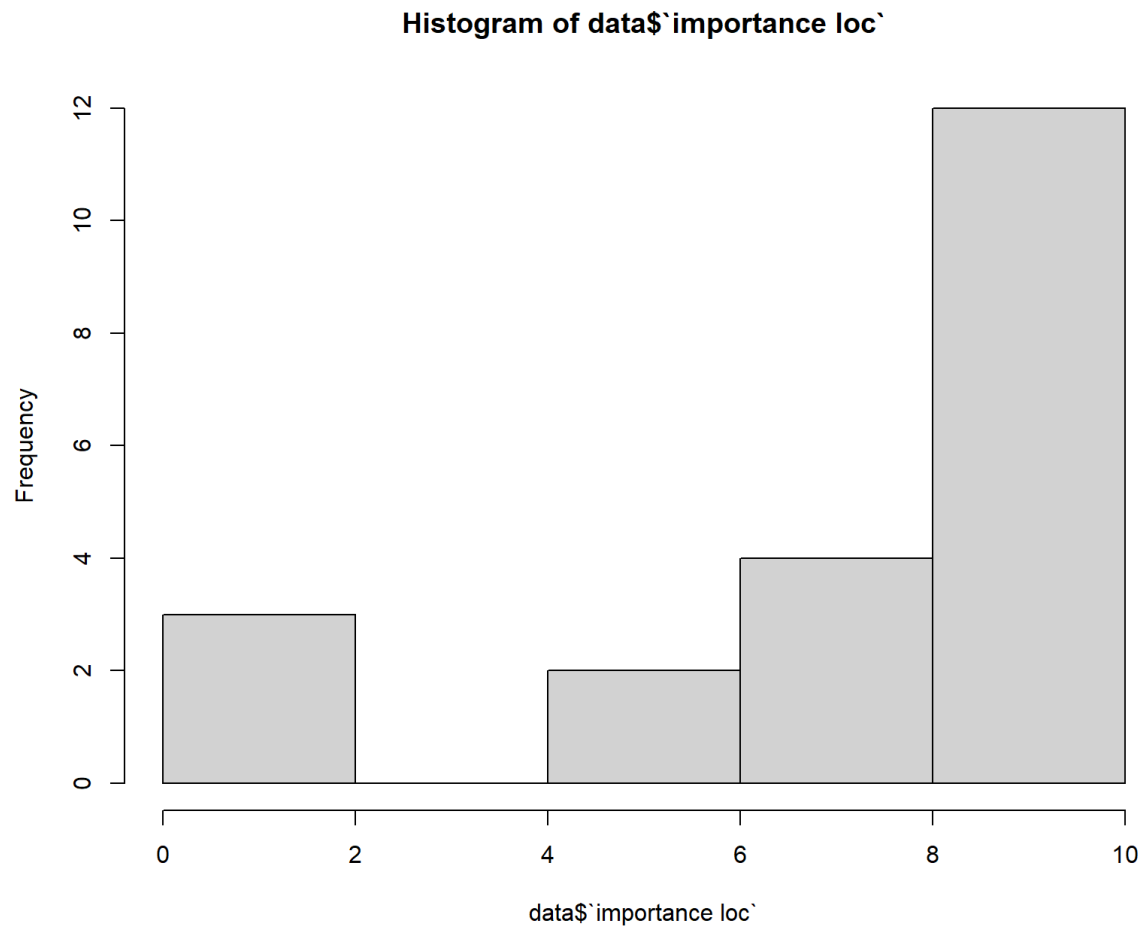
Graph 32 - *boxplot of importance of review*



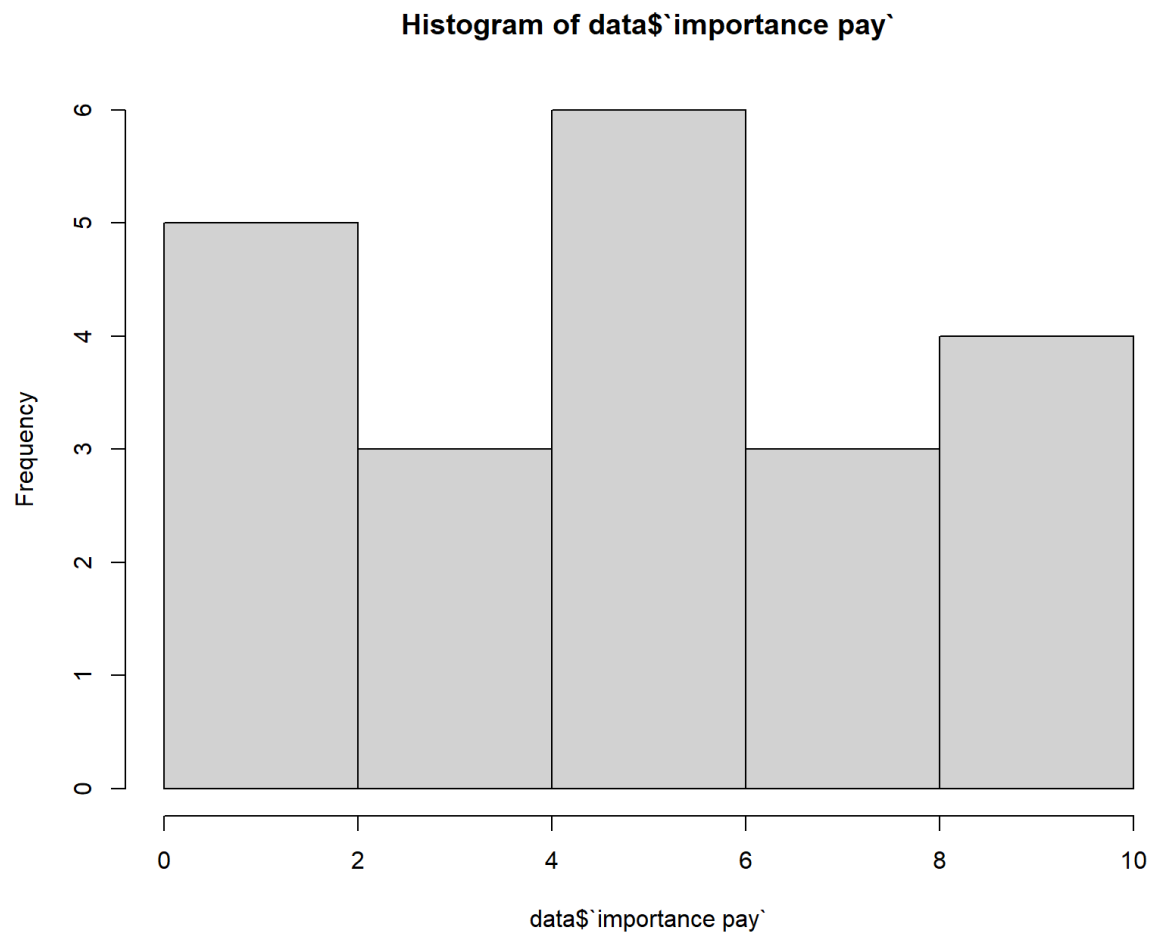
Graph 33 - *Boxplot Importance of Location*



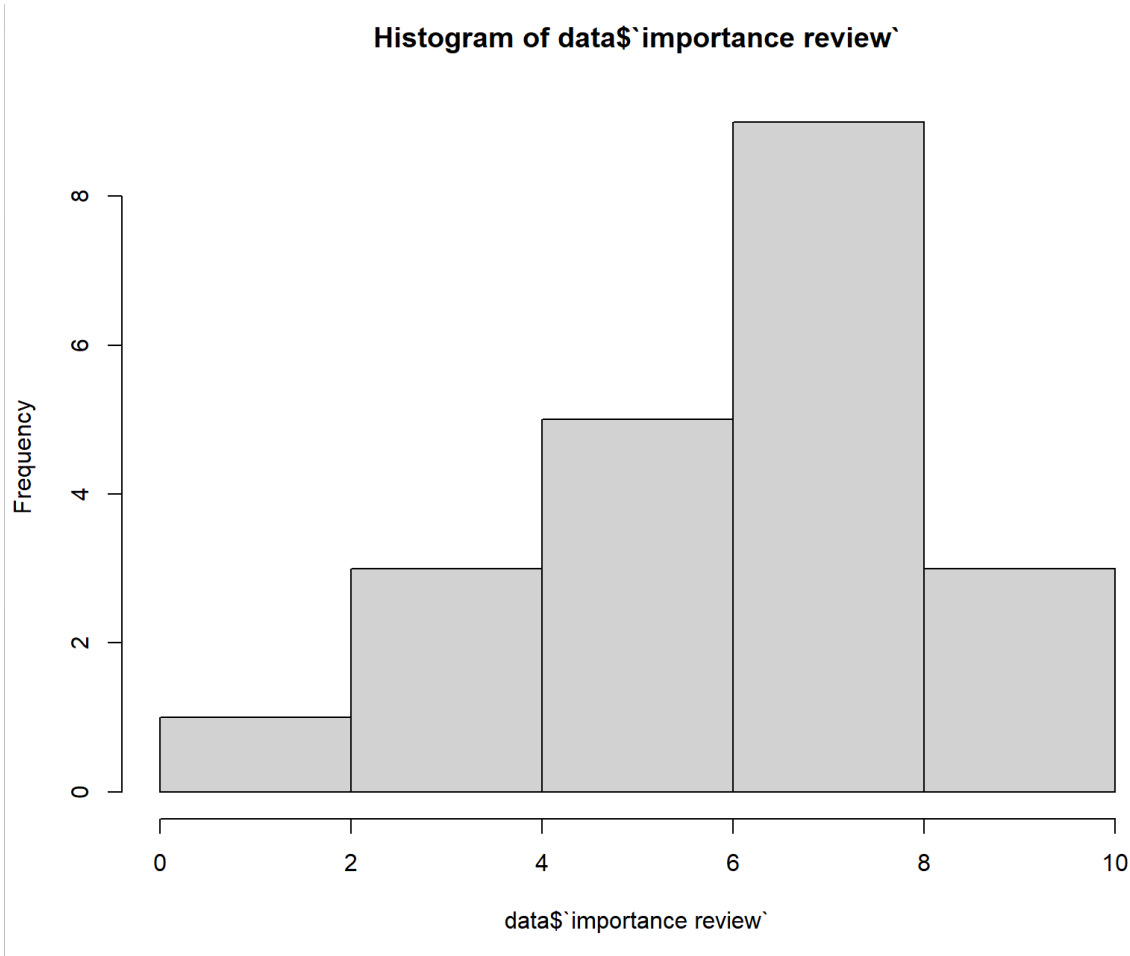
Graph 34 - *Boxplot of Importance of Pay*



Graph 35 - *Histogram of importance of location*



Graph 36 - *Histogram of Importance of Pay*



Graph 37 - Histogram of Importance of Review